

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 15:55:21 ; Search time 218.82 Seconds
(without alignments)
21.629 Million cell updates/sec

Title: US-09-396-196f-3
Perfect score: 25
Sequence: 1 gttgctgactcgtctgttgaagcg 25

Scoring table: OLIGO_MUC
Gapop 60.0 , Gapext 60.0

Searched: 324599 seqs, 94655562 residues

Word size : 9

Total number of hits satisfying chosen parameters: 12863

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_NA:*

- 1: /cgn2-6/ptodata/2/ina/5A.COMB.seq:*
- 2: /cgn2-6/ptodata/2/ina/5B.COMB.seq:*
- 3: /cgn2-6/ptodata/2/ina/6A.COMB.seq:*
- 4: /cgn2-6/ptodata/2/ina/6B.COMB.seq:*
- 5: /cgn2-6/ptodata/2/ina/PCITUS.COMB.seq:*
- 6: /cgn2-6/ptodata/2/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	25	100.0	1041	2	US-08-401-068-7
2	25	100.0	1041	2	US-08-846-338-7
3	25	100.0	5872	3	US-08-411-768B-1
4	25	100.0	5872	3	US-08-411-768B-6
5	16	64.0	5000	3	US-09-104-070-1
6	15	60.0	708	4	US-08-998-416-997
7	14	56.0	384	2	US-08-820-170A-8
8	14	56.0	384	2	US-09-055-699-8
9	14	56.0	384	4	US-09-273-565-8
10	14	56.0	735	5	PCT-US94-14106-54
11	14	56.0	1493	2	US-08-820-170A-9
12	14	56.0	1493	3	US-09-055-699-9
13	14	56.0	1493	4	US-09-273-565-9
14	14	56.0	1848	1	US-08-447-422-15
15	14	56.0	3243	2	US-08-339-276-6
16	14	56.0	3243	2	US-08-468-579B-6
17	14	56.0	3243	3	US-08-468-577B-6
18	14	56.0	3282	1	US-08-072-574-11
19	14	56.0	3282	1	US-08-486-270-11
20	14	56.0	3282	3	US-08-367-264-11
21	14	56.0	4085	1	US-08-072-574-7
22	14	56.0	4085	1	US-08-486-270-7
23	14	56.0	4085	3	US-08-367-264-7
24	14	56.0	4181	1	US-08-072-574-9
25	14	56.0	4181	3	US-08-486-270-9
26	14	56.0	4181	3	US-08-367-264-9
27	14	56.0	4207	4	US-08-660-148-1

C 28	14	56.0	4207	4	US-08-660-148-3	Sequence 3, Appl
C 29	14	56.0	4303	4	US-08-660-148-4	Sequence 4, Appl
C 30	14	56.0	4303	4	US-08-660-148-6	Sequence 6, Appl
C 31	14	56.0	4566	2	US-08-465-976A-1	Sequence 1, Appl
C 32	14	56.0	4566	2	US-08-982-412-1	Sequence 1, Appl
C 33	14	56.0	5319	1	US-08-169-927-1	Sequence 1, Appl
C 34	14	56.0	7326	1	US-08-194-468-1	Sequence 1, Appl
C 35	14	56.0	7344	3	US-08-961-739-1	Sequence 1, Appl
C 36	14	56.0	12284	2	US-08-876-991-1	Sequence 1, Appl
C 37	14	56.0	12284	2	US-09-059-853-1	Sequence 1, Appl
C 38	14	56.0	12804	2	US-08-592-874-1	Sequence 2, Appl
C 39	14	56.0	28804	3	US-09-096-942-2	Sequence 2, Appl
C 40	14	56.0	28804	3	US-09-096-867-2	Sequence 2, Appl
C 41	13	52.0	749	4	US-08-998-416-727	Sequence 727, App
C 42	13	52.0	1161	5	US-08-153-848-31	Sequence 31, Appl
C 43	13	52.0	1161	5	PCT-US93-11153-31	Sequence 31, Appl
C 44	13	52.0	1242	1	US-08-252-966B-13	Sequence 13, Appl
C 45	13	52.0	1729	3	US-09-188-930-68	Sequence 68, Appl

ALIGNMENTS

RESULT 1
US-08-401-068-7
Sequence 7, Application US/08401068
Patent No. 5859335
GENERAL INFORMATION:
APPLICANT: Patton, David
TITLE OF INVENTION: Enhanced Biotin Biosynthesis in Plant Tissue
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESSES:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.308
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/401,068
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/351,970
FILING DATE: 08-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8614
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1041 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1038
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /product= "biotin synthase"
OTHER INFORMATION: /evidence= EXPERIMENTAL
US-08-401-068-7


```
LOCATION: 5098..5574
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start= 5098
OTHER INFORMATION: /function= "unknown", involved in biotin synthesis"
OTHER INFORMATION: /product= "protein"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "ORF1"
OTHER INFORMATION: /number= 6
FEATURE:
NAME/KEY: -10_signal
LOCATION: 45..49
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /standard_name= "promoter ptac"
FEATURE:
NAME/KEY: -35_signal
LOCATION: 23..28
OTHER INFORMATION: /standard_name= "promoter ptac"
FEATURE:
NAME/KEY: RBS
LOCATION: 105..119
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /standard_name= "biob RBS no.9"
FEATURE:
NAME/KEY: RBS
LOCATION: 2284..2297
OTHER INFORMATION: /standard_name= "bioc RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 3742..3752
OTHER INFORMATION: /standard_name= "bioc RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 5088..5100
OTHER INFORMATION: /standard_name= "ORF1 RBS"
FEATURE:
NAME/KEY: terminator
LOCATION: 5583..5644
OTHER INFORMATION: /standard_name= "rho-independent
OTHER INFORMATION: transcriptional terminator"
FEATURE:
NAME/KEY: stem_loop
LOCATION: 5583..5605
FEATURE:
NAME/KEY: promoter
LOCATION: 1..96
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /function= "promoter ptac"
OTHER INFORMATION: /evidence= EXPERIMENTAL
PUBLICATION INFORMATION:
DOCUMENT NUMBER: WO 87/01391 B1
FILING DATE: 26-AUG-1986
PUBLICATION DATE: 07-APR-1993
US-08-411-768B-1

Query Match 100.0%; Score 25; DB 3; Length 5872;
Best Local Similarity 100.0%; Pred. No. 2.3e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
RESULT 4
US-08-411-768B-6
; Sequence 6, Application US/08411768B
; Patent No. 6083712
; GENERAL INFORMATION:
; APPLICANT: Olwen Birch
; APPLICANT: Johann Brass
```

```
OY 1 gtctgcatctgtgtttgaagc 25
Db 173 GTTCTGATCTGCTGTTGAAGCG 197
```

```
APPLICANT: Martin Fuhrmann
APPLICANT: Nicholas Shaw
TITLE OF INVENTION: Biotechnological Method
TITLE OF INVENTION: of Producing Biotin
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Wordperfect
SOFTWARE: Version 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/411,768B
FILING DATE: 31-March-95
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 3124/92
FILING DATE: 02-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 2134/93
FILING DATE: 15-JUL-1993
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 5872 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
STRAIN: DSM498
IMMEDIATE SOURCE:
CLONE: pBO30A15-9
FEATURE:
NAME/KEY: CDS
LOCATION: 1154..2308
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start= 1154
OTHER INFORMATION: /ec_number= 2.3.1.47
OTHER INFORMATION: /product= "KAPA synthase"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "biob"
OTHER INFORMATION: /number= 2
OTHER INFORMATION: /standard_name= "8-Amino-7-oxononanoate synthase"
FEATURE:
NAME/KEY: CDS
LOCATION: 3043..3753
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start= 3043
OTHER INFORMATION: /ec_number= 6.3.3.3
OTHER INFORMATION: /product= "DTB synthase"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "biob"
OTHER INFORMATION: /number= 4
OTHER INFORMATION: /standard_name= "Dethiobiotin synthase"
FEATURE:
NAME/KEY: RBS
LOCATION: 1141..1156
OTHER INFORMATION: /standard_name= "biob RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 3030..3045
OTHER INFORMATION: /standard_name= "biob RBS"
PUBLICATION INFORMATION:
DOCUMENT NUMBER: WO 87/01391 B1
```

;; FILING DATE: 26-AUG-1986
;; PUBLICATION DATE: 07-APR-1993
US-08-411-768B-6

Query Match 100.0%; Score 25; DB 3; Length 5872;
Best Local Similarity 100.0%; Pred. No. 2.3e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gtctgcatctgctgtttgaagc 25
|||||
Db 173 GTTGTGATCTGCTGTTGAAGC 197

RESULT 5

US-09-104-070-1
; Sequence 1, Application US/09104070
; Patent No. 6153741
; GENERAL INFORMATION:
; APPLICANT: Richards, Eric J.
; TITLE OF INVENTION: DNA Methylation Gene from Plants
; FILE REFERENCE: Wash U CI-0014 US
; CURRENT APPLICATION NUMBER: US/09/104, 070
; EARLIER FILING DATE: 1998-06-24
; EARLIER APPLICATION NUMBER: US 60/083,612
; NUMBER OF SEQ. ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ. ID NO 1
; LENGTH: 5000
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (782)...(1252)
; NAME/KEY: CDS
; LOCATION: (1354)...(1440)
; NAME/KEY: CDS
; LOCATION: (1549)...(1895)
; NAME/KEY: CDS
; LOCATION: (1976)...(2165)
; NAME/KEY: CDS
; LOCATION: (2251)...(2437)
; NAME/KEY: CDS
; LOCATION: (2559)...(2629)
; NAME/KEY: CDS
; LOCATION: (2703)...(2892)
; NAME/KEY: CDS
; LOCATION: (2975)...(3070)
; NAME/KEY: CDS
; LOCATION: (3148)...(3242)
; NAME/KEY: CDS
; LOCATION: (3317)...(3436)
; NAME/KEY: CDS
; LOCATION: (3540)...(3659)
; NAME/KEY: CDS
; LOCATION: (3745)...(3843)
; NAME/KEY: CDS
; LOCATION: (3934)...(4038)
; NAME/KEY: CDS
; LOCATION: (4130)...(4354)
; NAME/KEY: CDS
; LOCATION: (4826)...(4755)
; OTHER INFORMATION: t-RNA-glu coding region
; FEATURE:
; NAME/KEY: mutation
; LOCATION: (785)...(786)
; OTHER INFORMATION: some rearrangement, deletion of G at 785 and
; FEATURE: insertion of 83 bp between 785 and 786
; NAME/KEY: mutation
; LOCATION: (3243)...(3243)

;; OTHER INFORMATION: ddm1-2 base pair substitution of G to A
US-09-104-070-1

Query Match 64.0%; Score 16; DB 3; Length 5000;
Best Local Similarity 100.0%; Pred. No. 1.7;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 tgcggatctgctgt 18
|||||
Db 1132 tgcggatctgctgt 1147

RESULT 6

US-08-998-416-997/c
; Sequence 997, Application US/08998416
; Patent No. 6239264
; GENERAL INFORMATION:
; APPLICANT: Philippson, Peter
; APPLICANT: Pohlmann, Rainer
; APPLICANT: Steiner, Sabine
; APPLICANT: Mohr, Christine
; APPLICANT: Wendland, Jürgen
; APPLICANT: Knechtle, Philipp
; APPLICANT: Redischung, Corinne
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSTYPTI
; NUMBER OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 1152
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6239264artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: No. 6239264th Carolina
; COUNTRY: USA
; ZIP: 27709
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/998,416
; FILING DATE: 24-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 0016/97
; FILING DATE: 31-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/GC1976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 997:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 708 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: PAC1612RP
US-08-998-416-997

Query Match 60.0%; Score 15; DB 4; Length 708;
Best Local Similarity 100.0%; Pred. No. 5.7;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 tgcggatctgctgt 17
|||||
Db 470 TCTGTGATCTGCTGT 456

```

RESULT 7
US-08-820-170A-8/C
; Sequence 8, Application US/08820170A
; Patent No. 5831058
; GENERAL INFORMATION:
; APPLICANT: Tsutomu, FUJIWARA
; APPLICANT: Takeshi, WATANABE
; APPLICANT: Masato, HORIE
; APPLICANT: Toyomasa, KATAGIRI
; TITLE OF INVENTION: HUMAN GENE
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas
; STREET: 2100 Pennsylvania Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States
; ZIP: 20037-3202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/820,170A
; FILING DATE:
; CLASSIFICATION: 536
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 293-7060
; TELEFAX: (202) 293-7860
; TELEX: 6491103
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 384 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA(genomic)
; US-08-820-170A-8

Query Match          56.0%; Score 14; DB 2; Length 384;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 tggatctgctgttt 19
    |||
Db 350 TGGATCTGCTGTTT 337

RESULT 8
US-09-055-699-8/C
; Sequence 8, Application US/09055699
; Patent No. 6005088
; GENERAL INFORMATION:
; APPLICANT: Tsutomu, FUJIWARA
; APPLICANT: Takeshi, WATANABE
; APPLICANT: Masato, HORIE
; APPLICANT: Toyomasa, KATAGIRI
; TITLE OF INVENTION: HUMAN GENE
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas
; STREET: 2100 Pennsylvania Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States
; ZIP: 20037-3202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/055,699
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/820,170
; FILING DATE:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 293-7060
; TELEFAX: (202) 293-7860
; TELEX: 6491103
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 384 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA(genomic)
; US-09-055-699-8

Query Match          56.0%; Score 14; DB 3; Length 384;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 6 tggatctgctgttt 19
    |||
Db 350 TGGATCTGCTGTTT 337

```

```

RESULT 9
US-09-273-565-8/C
; Sequence 8, Application US/09273565A
; Patent No. 6166190
; GENERAL INFORMATION:
; APPLICANT: FUJIWARA, TSUTOMU
; APPLICANT: WATANABE, TAKESHI
; APPLICANT: HORIE, MASATO
; TITLE OF INVENTION: AN ISOLATED NUCLEIC ACID MOLECULE ENCODING HUMAN
; SKELETAL MUSCLE-SPECIFIC UBIQUITIN-CONJUGATING ENZYME
; FILE REFERENCE: 0-53599
; CURRENT APPLICATION NUMBER: US/09/273,565A
; CURRENT FILING DATE: 1999-03-22
; EARLIER APPLICATION NUMBER: 09/055,699
; EARLIER FILING DATE: 1998-04-07
; EARLIER APPLICATION NUMBER: 08/820,170
; EARLIER FILING DATE: 1997-03-19
; EARLIER APPLICATION NUMBER: JP 63410/1996
; EARLIER FILING DATE: 1996-03-19
; EARLIER APPLICATION NUMBER: JP 69163/1997
; EARLIER FILING DATE: 1997-03-05
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 8
; LENGTH: 384
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-273-565-8

```

```

Query Match          56.0%; Score 14; DB 4; Length 384;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 6 tggatctgctgttt 19
    |||
Db 350 TGGATCTGCTGTTT 337

```

```

RESULT 10
PCT-US94-14106-54

```

```
; Sequence 54, Application PC/TUS9414106
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Process for Generating Specific Antibodies
; NUMBER OF SEQUENCES: 61
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII (text)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/14106
; FILING DATE:
; CLASSIFICATION:
; INFORMATION FOR SEQ ID NO: 54:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 735 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 67..735
; PCT-US94-14106-54

Query Match          56.0%; Score 14; DB 5; Length 735;
Best Local Similarity 100.0%; Pred No. 20;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 tctgtatctgctg 16
Db 465 TCGTGATCTGCTG 478

RESULT 11
US-08-820-170A-9/c
; Sequence 9, Application US/08820170A
; Patent No. 5831058
; GENERAL INFORMATION:
; APPLICANT: Tsutomu, FUJIMARA
; APPLICANT: Takeshi, MATANABE
; APPLICANT: Masato, HORIE
; APPLICANT: Toyomasa, KATAGIRI
; TITLE OF INVENTION: HUMAN GENE
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas
; STREET: 2100 Pennsylvania Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States
; ZIP: 20037-3202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/820,170A
; FILING DATE:
; CLASSIFICATION: 536
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 293-7060
; TELEFAX: (202) 293-7860
; TELEX: 6491103
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1493 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
```

```
; MOLECULE TYPE: DNA(genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; IMMEDIATE SOURCE:
; LIBRARY: Human fetal brain cDNA library
; CLONE: GEN-025F07
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 95..478
; US-08-820-170A-9
```

```
Query Match          56.0%; Score 14; DB 2; Length 1493;
Best Local Similarity 100.0%; Pred No. 20;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 tggatctgctgtt 19
Db 444 TGGATCTGCTGTTT 431
```

```
RESULT 12
US-09-055-699-9/c
; Sequence 9, Application US/09055699
; Patent No. 6005088
; GENERAL INFORMATION:
; APPLICANT: Tsutomu, FUJIMARA
; APPLICANT: Takeshi, MATANABE
; APPLICANT: Masato, HORIE
; APPLICANT: Toyomasa, KATAGIRI
; TITLE OF INVENTION: HUMAN GENE
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas
; STREET: 2100 Pennsylvania Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States
; ZIP: 20037-3202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/055,699
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/820,170
; FILING DATE:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 293-7060
; TELEFAX: (202) 293-7860
; TELEX: 6491103
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1493 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA(genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; IMMEDIATE SOURCE:
; LIBRARY: Human fetal brain cDNA library
; CLONE: GEN-025F07
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 95..478
; US-09-055-699-9
```

Query Match 56.0%; Score 14; DB 3; Length 1493;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 tggatctgctgtt 19
 |||||
 Db 444 TGGATCTGCTGTT 431

RESULT 13
 US-09-273-565-9/c
 ; Sequence 9, Application US/09273565A
 ; Patent No. 6166190
 ; GENERAL INFORMATION:
 ; APPLICANT: FUJIMURA, TSUTOMU
 ; APPLICANT: MATANABE, TAKESHI
 ; APPLICANT: HORIE, MASATO
 ; TITLE OF INVENTION: AN ISOLATED NUCLEIC ACID MOLECULE ENCODING HUMAN
 ; TITLE OF INVENTION: SKELETAL MUSCLE-SPECIFIC UBIQUITIN-CONJUGATING ENZYME
 ; FILE REFERENCE: Q-53599
 ; CURRENT APPLICATION NUMBER: US/09/273,565A
 ; EARLIER FILING DATE: 1999-03-22
 ; EARLIER APPLICATION NUMBER: 09/055,699
 ; EARLIER FILING DATE: 1998-04-07
 ; EARLIER APPLICATION NUMBER: 08/820,170
 ; EARLIER FILING DATE: 1997-03-19
 ; EARLIER APPLICATION NUMBER: JP 63410/1996
 ; EARLIER FILING DATE: 1996-03-19
 ; EARLIER APPLICATION NUMBER: JP 69163/1997
 ; EARLIER FILING DATE: 1997-03-05
 ; NUMBER OF SEQ. ID NOS: 95
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 9
 ; LENGTH: 1493
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (95)..(478)
 ; US-09-273-565-9

Query Match 56.0%; Score 14; DB 4; Length 1493;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 tggatctgctgtt 19
 |||||
 Db 444 TGGATCTGCTGTT 431

RESULT 14
 US-08-447-422-15/c
 ; Sequence 15, Application US/08447422
 ; Patent No. 5686579
 ; GENERAL INFORMATION:
 ; APPLICANT: SHAMI, Ezekiel Y.
 ; APPLICANT: ROTHSTEIN, Aser
 ; APPLICANT: RAMJESINGH, Mohanb
 ; TITLE OF INVENTION: Use of Antibody/Antigen Interactions To
 ; TITLE OF INVENTION: Protect or Modulate Biological Activity
 ; NUMBER OF SEQUENCES: 15
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Foley & Lardner
 ; STREET: 3000 K Street, N.W., Suite 500
 ; CITY: Washington, D.C.
 ; COUNTRY: USA
 ; ZIP: 20007-5109
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/447,422
 ; FILING DATE:
 ; CLASSIFICATION: 530
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/081,410
 ; FILING DATE: 22-JUN-1993
 ; APPLICATION NUMBER: US 07/938,505
 ; FILING DATE: 31-AUG-1992
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/205,748
 ; FILING DATE: 21-JUN-1988
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: BENT, Stephen A.
 ; REGISTRATION NUMBER: 29,768
 ; REFERENCE/DOCKET NUMBER: 17923/102 HYLL
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (202)672-5300
 ; TELEFAX: (202)672-5399
 ; TELEX: 904136
 ; INFORMATION FOR SEQ. ID NO: 15:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1848 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; US-08-447-422-15

Query Match 56.0%; Score 14; DB 1; Length 1848;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 tgcgtgctgctg 16
 |||||
 Db 1829 TCGTGTGCTGCTG 1816

RESULT 15
 US-08-239-276-6/c
 ; Sequence 6, Application US/08239276
 ; Patent No. 5840836
 ; GENERAL INFORMATION:
 ; APPLICANT: Rabin, Daniel
 ; TITLE OF INVENTION: PANCREATIC ISLET CELL ANTIGENS
 ; TITLE OF INVENTION: OBTAINED BY MOLECULAR CLONING
 ; NUMBER OF SEQUENCES: 10
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Sprung Kramer Schaefer & Briscoe
 ; STREET: 660 White Plains Road
 ; CITY: Tarrytown
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10591-5144
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage
 ; COMPUTER: Apple Macintosh
 ; OPERATING SYSTEM: System 7.5
 ; SOFTWARE: WordPerfect
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/239,276
 ; FILING DATE: 05-MAY-1994
 ; CLASSIFICATION: 530
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/872,646
 ; FILING DATE: 08-JUN-1992
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/715,181
 ; FILING DATE: 14-JUN-1991
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/441,703
 ; FILING DATE: 04-DEC-1989
 ; PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/312,543
FILING DATE: 17-FEB-1989
ATTORNEY/AGENT INFORMATION:
NAME: Kurt G. Briscoe
REGISTRATION NUMBER: 33,141
REFERENCE/DOCKET NUMBER: MDI 251.4-KGB
TELECOMMUNICATION INFORMATION:
TELEPHONE: (914) 332-1700
TELEFAX: (914) 332-1844
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 3243 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-239-276-6

Query Match 56.0%; Score 14; DB 2; Length 3243;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 tgcgtgactctgctg 16
|||||
DB 413 TGCTGGATCTGCTG 400

RESULT 16
US-08-468-579B-6/c
Sequence 6, Application US/08468579B
Patent No. 5981700
GENERAL INFORMATION:
APPLICANT: Rabin, Daniel
TITLE OF INVENTION: PANCREATIC ISLET CELL ANTIGENS
NUMBER OF INVENTION: OBTAINED BY MOLECULAR CLONING
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sprung Kramer Schaefer & Briscoe
STREET: 660 White Plains Road
CITY: Tarrytown
STATE: New York
COUNTRY: USA
ZIP: 10591-5144
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage
COMPUTER: Apple Macintosh
OPERATING SYSTEM: System 7.5
SOFTWARE: WordPerfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/468,579B
FILING DATE: 06-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/239,276
FILING DATE: 05-MAY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/872,646
FILING DATE: 08-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/715,181
FILING DATE: 14-JUN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/441,703
FILING DATE: 04-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/312,543
FILING DATE: 17-FEB-1989
ATTORNEY/AGENT INFORMATION:
NAME: Kurt G. Briscoe
REGISTRATION NUMBER: 33,141
REFERENCE/DOCKET NUMBER: MDI 251.5-KGB
TELECOMMUNICATION INFORMATION:
TELEPHONE: (914) 332-1700

TELEFAX: (914) 332-1844
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 3243 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-468-579B-6

Query Match 56.0%; Score 14; DB 2; Length 3243;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 tgcgtgactctgctg 16
|||||
DB 413 TGCTGGATCTGCTG 400

RESULT 17
US-08-468-577B-6/c
Sequence 6, Application US/08468577B
Patent No. 6001804
GENERAL INFORMATION:
APPLICANT: Rabin, Daniel
TITLE OF INVENTION: PANCREATIC ISLET CELL ANTIGENS
NUMBER OF INVENTION: OBTAINED BY MOLECULAR CLONING
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sprung Kramer Schaefer & Briscoe
STREET: 660 White Plains Road
CITY: Tarrytown
STATE: New York
COUNTRY: USA
ZIP: 10591-5144
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage
COMPUTER: Apple Macintosh
OPERATING SYSTEM: System 7.5
SOFTWARE: WordPerfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/468,577B
FILING DATE: 06-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/239,276
FILING DATE: 05-MAY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/872,646
FILING DATE: 08-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/715,181
FILING DATE: 14-JUN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/441,703
FILING DATE: 04-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/312,543
FILING DATE: 17-FEB-1989
ATTORNEY/AGENT INFORMATION:
NAME: Kurt G. Briscoe
REGISTRATION NUMBER: 33,141
REFERENCE/DOCKET NUMBER: MDI 251.8-KGB
TELECOMMUNICATION INFORMATION:
TELEPHONE: (914) 332-1700
TELEFAX: (914) 332-1844
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 3243 nucleotides
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-468-577B-6

Query Match 56.0%; Score 14; DB 3; Length 3243;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 tgcgtgactgctg 16
|||||
Db 413 TGCTGATCTGCTG 400

RESULT 18
US-08-072-574-11/c
; Sequence 11, Application US/08072574
; Patent No. 5521297
; GENERAL INFORMATION:
; APPLICANT: Daggett, Lorrle
; APPLICANT: Ellis, Steven B.
; APPLICANT: Liaw, Chen
; APPLICANT: Pontsler, Aaron
; TITLE OF INVENTION: HUMAN METABOTROPIC GLUTAMATE RECEPTORS,
; TITLE OF INVENTION: NOCLETIC ACIDS ENCODING SAME AND USES THEREOF
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
; STREET: 444 South Flower Street, Suite 2000
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 00719
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/072,574
; FILING DATE: 19930604
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Reiter, Stephen E.
; REGISTRATION NUMBER: 31,192
; REFERENCE/DOCKET NUMBER: P41 9383
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 213-622-7700
; TELEFAX: 213-489-4210
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3282 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 370..3003
; OTHER INFORMATION: /product= "HUMAN MGLUR5C"
; OTHER INFORMATION: /note= "Variant of MGLUR5A with truncated 3' end."
US-08-072-574-11

Query Match 56.0%; Score 14; DB 1; Length 3282;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 tgcgtgactgctg 16
|||||
Db 2950 TGCTGATCTGCTG 2937

RESULT 19
US-08-486-270-11/c
; Sequence 11, Application US/08486270

; Patent No. 5807689
; GENERAL INFORMATION:
; APPLICANT: Daggett, Lorrle
; APPLICANT: Ellis, Steven B.
; APPLICANT: Liaw, Chen
; APPLICANT: Pontsler, Aaron
; APPLICANT: Johnson, Edwin C.
; APPLICANT: Hess, Stephen D.
; TITLE OF INVENTION: HUMAN METABOTROPIC GLUTAMATE RECEPTORS,
; TITLE OF INVENTION: NOCLETIC ACIDS ENCODING SAME AND USES THEREOF
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
; STREET: 444 South Flower Street, Suite 2000
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/486,270
; FILING DATE: 02-JUN-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/072,574
; FILING DATE: 04-JUN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Reiter, Stephen E.
; REGISTRATION NUMBER: 31,192
; REFERENCE/DOCKET NUMBER: P41 9772
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-546-4737
; TELEFAX: 619-546-9392
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3282 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: both
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 370..3003
; OTHER INFORMATION: /product= "HUMAN MGLUR5C"
; OTHER INFORMATION: /note= "Variant of MGLUR5A with truncated 3' end."
US-08-486-270-11

Query Match 56.0%; Score 14; DB 1; Length 3282;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 tgcgtgactgctg 16
|||||
Db 2950 TGCTGATCTGCTG 2937

RESULT 20
US-08-367-264-11/c
; Sequence 11, Application US/08367264
; Patent No. 6001581
; GENERAL INFORMATION:
; APPLICANT: Daggett, Lorrle
; APPLICANT: Ellis, Steven B.
; APPLICANT: Liaw, Chen
; APPLICANT: Pontsler, Aaron
; APPLICANT: Johnson, Edwin C.
; APPLICANT: Hess, Stephen D.
; TITLE OF INVENTION: HUMAN METABOTROPIC GLUTAMATE RECEPTORS,

Query Match	56.0%	Score 14	DB 1	Length 4085
Best Local Similarity	100.0%	Pred. No. 21		
Matches 14	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	3	tgcctgagctcctcg	16	
Db	2950	tgcctgagctcctcg	2937	

RESULT 22
 : Sequence 7, Application US/08486270
 : Patent No. 5807689
 : GENERAL INFORMATION:
 : APPLICANT: Daggett, Lorie
 : APPLICANT: Ellis, Steven B.
 : APPLICANT: Liaw, Chen
 : APPLICANT: Pontsler, Aaron
 : APPLICANT: Johnson, Edwin C.
 : APPLICANT: Hess, Stephen D.
 : TITLE OF INVENTION: HUMAN METABOTROPIC GLUTAMATE RECEPTORS,
 : TITLE OF INVENTION: NUCLEIC ACIDS ENCODING SAME AND USES THEREOF
 : NUMBER OF SEQUENCES: 13
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Pretly, Schroeder, Brueggemann & Clark
 : STREET: 444 South Flower Street, Suite 2000
 : CITY: Los Angeles
 : STATE: CA
 : COUNTRY: USA
 : ZIP: 90071
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: PatentIn Release #1.0, Version #1.25
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/486,270
 : FILING DATE: 02-JUN-1994
 : CLASSIFICATION: 435
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US 08/072,574
 : FILING DATE: 04-JUN-1993
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Reiter, Stephen E.
 : REGISTRATION NUMBER: 31,192

REFERENCE/DOCKET NUMBER: FP41 9772
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-546-4737
TELEFAX: 619-546-9392
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 4085 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 370..3912
OTHER INFORMATION: /product= "HUMAN MGLUR5A"
US-08-486-270-7

Query Match 56.0%; Score 14; DB 1; Length 4085;
Best local Similarity 100.0%; Pred. No. 21;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 tgctgacatcgtcgtg 16
|||||
DB 2950 TGCTGATCTGCTG 2937

RESULT 23
US-08-367-264-7/C
Sequence 7, Application US/08367264
Patent No. 6001581
GENERAL INFORMATION:
APPLICANT: Daggett, Lorie
APPLICANT: Ellis, Steven B.
APPLICANT: Liaw, Chen
APPLICANT: Pontsler, Aaron
APPLICANT: Johnson, Edwin C.
APPLICANT: Hess, Stephen D.
TITLE OF INVENTION: HUMAN METABOTROPIC GLUTAMATE RECEPTORS,
TITLE OF INVENTION: NUCLEIC ACIDS ENCODING SAME AND USES THEREOF
NUMBER OF SEQUENCES: 13
CURRENT APPLICATION DATA:
CORRESPONDENCE ADDRESS:
ADDRESS: Pretly, Schroeder, Brueggemann & Clark
STREET: 444 South Flower Street, Suite 2000
CITY: Los Angeles
STATE: CA
COUNTRY: USA
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/367,264
FILING DATE: 02-JUN-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/072,574
FILING DATE: 04-JUN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Reiter, Stephen E.
REGISTRATION NUMBER: 31,192
REFERENCE/DOCKET NUMBER: FP41 9772
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-546-4737
TELEFAX: 619-546-9392
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 4085 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both

MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 370..3912
OTHER INFORMATION: /product= "HUMAN MGLUR5A"
US-08-367-264-7

Query Match 56.0%; Score 14; DB 3; Length 4085;
Best local Similarity 100.0%; Pred. No. 21;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 tgctgacatcgtcgtg 16
|||||
DB 2950 TGCTGATCTGCTG 2937

RESULT 24
US-08-072-574-9/C
Sequence 9, Application US/08072574
Patent No. 5521297
GENERAL INFORMATION:
APPLICANT: Daggett, Lorie
APPLICANT: Ellis, Steven B.
APPLICANT: Liaw, Chen
APPLICANT: Pontsler, Aaron
TITLE OF INVENTION: HUMAN METABOTROPIC GLUTAMATE RECEPTORS,
TITLE OF INVENTION: NUCLEIC ACIDS ENCODING SAME AND USES THEREOF
NUMBER OF SEQUENCES: 13
CURRENT APPLICATION DATA:
CORRESPONDENCE ADDRESS:
ADDRESS: Pretly, Schroeder, Brueggemann & Clark
STREET: 444 South Flower Street, Suite 2000
CITY: Los Angeles
STATE: CA
COUNTRY: USA
ZIP: 00719
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/072,574
FILING DATE: 19930604
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Reiter, Stephen E.
REGISTRATION NUMBER: 31,192
REFERENCE/DOCKET NUMBER: P41 9383
TELECOMMUNICATION INFORMATION:
TELEPHONE: 213-622-7700
TELEFAX: 213-489-4210
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 4181 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 370..4008
OTHER INFORMATION: /note= "Variant of MGLUR5A with 96 base pair
insertion between nucleotides 2999 and 3000."
US-08-072-574-9

Query Match 56.0%; Score 14; DB 1; Length 4181;
Best local Similarity 100.0%; Pred. No. 21;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 tgctgacatcgtcgtg 16

Db 2950 TCCTGATCTGCTG 2937

RESULT 25

US-08-486-270-9/c
; Sequence 9, Application US/08486270
; Patent No. 5807689

GENERAL INFORMATION:

APPLICANT: Daggett, Lorrise
APPLICANT: Ellis, Steven B.

APPLICANT: Liaw, Chen

APPLICANT: Pontsler, Aaron

APPLICANT: Johnson, Edwin C.

APPLICANT: Hess, Stephen D.

TITLE OF INVENTION: HUMAN METABOTROPIC GLUTAMATE RECEPTORS,

TITLE OF INVENTION: NUCLEIC ACIDS ENCODING SAME AND USES THEREOF

NUMBER OF SEQUENCES: 13

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark

STREET: 444 South Flower Street, Suite 2000

CITY: Los Angeles

STATE: CA

COUNTRY: USA

ZIP: 90071

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/486,270

FILING DATE: 02-JUN-1994

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/072,574

FILING DATE: 04-JUN-1993

ATTORNEY/AGENT INFORMATION:

NAME: Reiter, Stephen E.

REGISTRATION NUMBER: 31,192

REFERENCE/DOCKET NUMBER: FP41 9772

TELEPHONE: 619-546-4737

TELEFAX: 619-546-9392

INFORMATION FOR SEQ ID NO: 9:

SEQUENCE CHARACTERISTICS:

LENGTH: 4181 base pairs

TYPE: nucleic acid

STRANDEDNESS: both

TOPOLOGY: both

MOLECULE TYPE: CDNA

FEATURE:

NAME/KEY: CDS

LOCATION: 370..4008

OTHER INFORMATION: /product= "HUMAN MGUR5B"

OTHER INFORMATION: /note= "Variant of MGUR5A with 96 base pair

OTHER INFORMATION: insertion between nucleotides 2998 and 2999."

US-08-486-270-9

Query Match

Best Local Similarity 56.0%; Score 14; DB 1; Length 4181;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 2950 TCCTGATCTGCTG 2937

Y 3 tctgtgactctgctg 16

|||||

RESULT 26

US-08-367-264-9/c

; Sequence 9, Application US/08367264

; Patent No. 6001581

GENERAL INFORMATION:

APPLICANT: Daggett, Lorrise

APPLICANT: Ellis, Steven B.

APPLICANT: Liaw, Chen

APPLICANT: Pontsler, Aaron

APPLICANT: Johnson, Edwin C.

APPLICANT: Hess, Stephen D.

TITLE OF INVENTION: HUMAN METABOTROPIC GLUTAMATE RECEPTORS,

TITLE OF INVENTION: NUCLEIC ACIDS ENCODING SAME AND USES THEREOF

NUMBER OF SEQUENCES: 13

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark

STREET: 444 South Flower Street, Suite 2000

CITY: Los Angeles

STATE: CA

COUNTRY: USA

ZIP: 90071

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/367,264

FILING DATE: 02-JUN-1994

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/072,574

FILING DATE: 04-JUN-1993

ATTORNEY/AGENT INFORMATION:

NAME: Reiter, Stephen E.

REGISTRATION NUMBER: 31,192

REFERENCE/DOCKET NUMBER: FP41 9772

TELEPHONE: 619-546-4737

TELEFAX: 619-546-9392

INFORMATION FOR SEQ ID NO: 9:

SEQUENCE CHARACTERISTICS:

LENGTH: 4181 base pairs

TYPE: nucleic acid

STRANDEDNESS: both

TOPOLOGY: both

MOLECULE TYPE: CDNA

FEATURE:

NAME/KEY: CDS

LOCATION: 370..4008

OTHER INFORMATION: /product= "HUMAN MGUR5B"

OTHER INFORMATION: /note= "Variant of MGUR5A with 96 base pair

OTHER INFORMATION: insertion between nucleotides 2998 and 2999."

US-08-367-264-9

Query Match

Best Local Similarity 56.0%; Score 14; DB 3; Length 4181;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 2950 TCCTGATCTGCTG 2937

Y 3 tctgtgactctgctg 16

|||||

RESULT 27

US-08-660-148-1/c

; Sequence 1, Application US/08660148

; Patent No. 6211353

GENERAL INFORMATION:

APPLICANT: Burnett, J. P.

APPLICANT: Mayne, Nancy G.

APPLICANT: Sharp, Robert L.

APPLICANT: Snyder, Yvonne M.

TITLE OF INVENTION: EXCITATORY AMINO RECEPTOR PROTEIN AND

TITLE OF INVENTION: RELATED NUCLEIC ACID COMPOUNDS

NUMBER OF SEQUENCES: 6

;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Eli Lilly and Company
;; STREET: Lilly Corporate Center
;; CITY: Indianapolis
;; STATE: Indiana
;; COUNTRY: United States of America
;; ZIP: 46285
;;
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/660,148
;;
;; CLASSIFICATION:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US/08/282,853
;; FILING DATE:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Gaylo, Paul J.
;; REGISTRATION NUMBER: 36,808
;; REFERENCE/DOCKET NUMBER: X-9419
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (317) 276-0756
;; TELEFAX: (317) 276-3861
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 4207 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: CDNA
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 460..3999
;;
;; US-08-660-148-1

Query Match 56.0%; Score 14; DB 4; Length 4207;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 tctgtgactctgctg 16
|||||

Db 3040 TCGCTGATCTGCTG 3027

RESULT 28
US-08-660-148-3/C
; Sequence 3, Application US/08660148
; Patent No. 6211353
; GENERAL INFORMATION:
; APPLICANT: Burnett, J. P.
; APPLICANT: Mayne, Nancy G.
; APPLICANT: Sharp, Robert L.
; APPLICANT: Snyder, Yvonne M.
; TITLE OF INVENTION: EXCITATORY AMINO RECEPTOR PROTEIN AND
; TITLE OF INVENTION: RELATED NUCLEIC ACID COMPOUNDS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: United States of America
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

;; APPLICATION NUMBER: US/08/660,148
;; FILING DATE:
;; CLASSIFICATION:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US/08/282,853
;; FILING DATE:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Gaylo, Paul J.
;; REGISTRATION NUMBER: 36,808
;; REFERENCE/DOCKET NUMBER: X-9419
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (317) 276-0756
;; TELEFAX: (317) 276-3861
;; INFORMATION FOR SEQ ID NO: 3:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 4207 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: mRNA
;;
;; US-08-660-148-3

Query Match 56.0%; Score 14; DB 4; Length 4207;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 tctgtgactctgctg 16
|||||

Db 3040 TCGCTGATCTGCTG 3027

RESULT 29
US-08-660-148-4/C
; Sequence 4, Application US/08660148
; Patent No. 6211353
; GENERAL INFORMATION:
; APPLICANT: Burnett, J. P.
; APPLICANT: Mayne, Nancy G.
; APPLICANT: Sharp, Robert L.
; APPLICANT: Snyder, Yvonne M.
; TITLE OF INVENTION: EXCITATORY AMINO RECEPTOR PROTEIN AND
; TITLE OF INVENTION: RELATED NUCLEIC ACID COMPOUNDS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: United States of America
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/660,148
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/282,853
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Gaylo, Paul J.
; REGISTRATION NUMBER: 36,808
; REFERENCE/DOCKET NUMBER: X-9419
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317) 276-0756
; TELEFAX: (317) 276-3861
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4303 base pairs

TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 460..4095
US-08-660-148-4

Query Match

Best Local Similarity 56.0%; Score 14; DB 4; Length 4303;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 tctgtatctgtcgt 16
|||||
DB 3040 TGCTGATCTGCTG 3027

RESULT 30

US-08-660-148-6/c
Sequence 6, Application US/08660148
Patent No. 6211353
GENERAL INFORMATION:
APPLICANT: Burnett, J. P.
APPLICANT: Mayne, Nancy G.
APPLICANT: Sharp, Robert L.
APPLICANT: Snyder, Yvonne M.
TITLE OF INVENTION: EXCITATORY AMINO RECEPTOR PROTEIN AND
NUMBER OF SEQUENCES: 6
RELATED NUCLEIC ACID COMPOUNDS
CORRESPONDENCE ADDRESS:
ADDRESSEE: Eli Lilly and Company
STREET: Lilly Corporate Center
CITY: Indianapolis
STATE: Indiana
COUNTRY: United States of America
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/660,148
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/282,853
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Gaylo, Paul J.
REGISTRATION NUMBER: 36,808
REFERENCE/DOCKET NUMBER: X-9419
TELECOMMUNICATION INFORMATION:
TELEPHONE: (317) 276-0756
TELEFAX: (317) 276-3861
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 4303 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: mRNA
US-08-660-148-6

Query Match

Best Local Similarity 56.0%; Score 14; DB 4; Length 4303;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 tctgtatctgtcgt 16
|||||

DB 3040 TGCTGATCTGCTG 3027

RESULT 31

US-08-465-976A-1/c
Sequence 1, Application US/08465976A
Patent No. 5869632
GENERAL INFORMATION:
APPLICANT: SOPPET, DANIEL R
APPLICANT: LI, YI
APPLICANT: ROSEN, CRAIG A
APPLICANT: RUBEN, STEVEN M
TITLE OF INVENTION: HUMAN G-PROTEIN RECEPTOR
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: CARELLA, BYRNE, BAIN GIFFILLAN, CECCHI
ADDRESSEE: STEWART & OLSTEIN
STREET: 6 BECKER FARM ROAD
CITY: ROSELAND
STATE: NJ
COUNTRY: US
ZIP: 07068
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/465,976A
FILING DATE: 06-JUN-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: FERRARO, GREGORY F
REGISTRATION NUMBER: 36,134
REFERENCE/DOCKET NUMBER: 325800-444
TELECOMMUNICATION INFORMATION:
TELEPHONE: (201) 994-1700
TELEFAX: (201) 994-1744
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4566 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 212..2863
US-08-465-976A-1

Query Match

Best Local Similarity 56.0%; Score 14; DB 2; Length 4566;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 ctgtatctgtcgt 18
|||||
DB 4182 CTGATCTGCTGTT 4169

RESULT 32

US-08-982-412-1/c
Sequence 1, Application US/08982412
Patent No. 5958729
GENERAL INFORMATION:
APPLICANT: SOPPET, DANIEL R
APPLICANT: LI, YI
APPLICANT: ROSEN, CRAIG A
APPLICANT: RUBEN, STEVEN M
TITLE OF INVENTION: HUMAN G-PROTEIN RECEPTOR
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: HUMAN GENOME SCIENCES, INC.

```

STREET: 9410 KEY WEST AVENUE
CITY: ROCKVILLE,
STATE: MD
COUNTRY: US
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/982,412
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: BROOKES, ANDERS A
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PF181PCT2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4566 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 212..2863
US-08-982-412-1

Query Match      56.0%; Score 14; DB 2; Length 4566;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 cttgactcgtcgtt 18
    |||
Db 4182 CTGCGATCTGCTGT 4169

RESULT 33
US-08-169-927-1
; Sequence 1, Application US/08169927
; Patent No. 5783441
; GENERAL INFORMATION:
; APPLICANT: Carl, Mitchell
; APPLICANT: Dobson, Michael E.
; APPLICANT: Ching, Gregory A.
; TITLE OF INVENTION: Gene and protein Applicable to the
; TITLE OF INVENTION: Preparation of vaccines for Rickettsia prowazekii and
; TITLE OF INVENTION: Rickettsia typhi and the Detection of Both
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Counsel, Naval Medical R & D Command
; STREET: Bldg. 1, T-12, 8901 Wisconsin Ave.
; CITY: Bethesda
; STATE: MD
; COUNTRY: USA
; ZIP: 20889-5606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/169,927
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:

```

```

APPLICATION NUMBER: US 07/742,128
FILING DATE: 08/09/91
ATTORNEY/AGENT INFORMATION:
NAME: Spevack, A. David
REGISTRATION NUMBER: 24,743
REFERENCE/DOCKET NUMBER: 75,976
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 295-6759
TELEFAX: (301) 295-1022
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5319 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: internal
ORIGINAL SOURCE:
ORGANISM: Rickettsia prowazekii
STRAIN: Breinl
FEATURE:
NAME/KEY: -35_signal
LOCATION: 340..345
FEATURE:
NAME/KEY: -10_signal
LOCATION: 363..368
FEATURE:
NAME/KEY: CDS
LOCATION: 391..5226
FEATURE:
NAME/KEY: RBS
LOCATION: 379..386
FEATURE:
NAME/KEY: stem_loop
LOCATION: 5270..5306
PUBLICATION INFORMATION:
AUTHORS: Carl, M.
AUTHORS: Dobson, M. E.
AUTHORS: Ching, W. M.
AUTHORS: Dasch, G. A.
TITLE: Characterization of the gene encoding the
TITLE: protective s-layer protein of Rickettsia
TITLE: prowazekii; presence of a truncated identical
TITLE: homolog in rickettsia typhi
JOURNAL: Proc. Natl. Acad. Sci. U.S.A.
DATE: 1990
RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 5319
US-08-169-927-1

Query Match      56.0%; Score 14; DB 1; Length 5319;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gttagcgtacgtc 14
    |||
Db 4321 GTTCTGATCTGC 4334

RESULT 34
US-08-194-468-1/c
; Sequence 1, Application US/08194468
; Patent No. 5750336
; GENERAL INFORMATION:
; APPLICANT: Montminy, Marc R.
; TITLE OF INVENTION: ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: COMPOUNDS WHICH INHIBIT ACTIVATION OF CAMP AND MITOGEN
; TITLE OF INVENTION: RESPONSIVE GENES
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark

```

STREET: 444 South Flower Street, Suite 2000
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90071

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/194,468
FILING DATE: 10-FEB-1994
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: Reiter, Stephen F.
REGISTRATION NUMBER: 31,192
REFERENCE/DOCKET NUMBER: P41 9672
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619)-546-9392
TELEFAX: (619)-546-9392
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 7326 base pairs
TYPE: nucleic acid
STRANDEDNESS: single

TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHEICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 1..7323
US-08-194-468-1

Query Match 56.0%; Score 14; DB 1; Length 7326;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 tgcgtgacatcgtcgt 16
|||||
Db 5573 TGCCTGATCTGCTG 5560

RESULT 35
US-08-961-739-1/C
Sequence 1, Application US/08961739A
PATENT NO. 6083583
GENERAL INFORMATION:
APPLICANT: Montminy, Marc R.
TITLE OF INVENTION: Methods for Treating Diabetes Mellitus
FILE REFERENCE: SALK1650-1
CURRENT APPLICATION NUMBER: US/08/961,739A
CURRENT FILING DATE: 1997-10-31
EARLIER FILING DATE: 1994-02-10
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 7326
TYPE: DNA
ORGANISM: Mus
FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(7326)
NAME/KEY: misc.feature
LOCATION: (1)...(7326)
OTHER INFORMATION: n = A,T,C or G
US-08-961-739-1

Query Match 56.0%; Score 14; DB 3; Length 7344;

Best Local Similarity 100.0%; Pred. No. 21;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 tgcgtgacatcgtcgt 16
|||||
Db 5573 TGCCTGATCTGCTG 5560

RESULT 36
US-08-876-991-1
Sequence 1, Application US/08876991
PATENT NO. 5925360
GENERAL INFORMATION:
APPLICANT: Gregor Meyers, Tillmann R menapf,
APPLICANT: Heinz-J. Igen Thiel
TITLE OF INVENTION: Hog cholera virus vaccine and diagnostic
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Organon Teknika Corporation
STREET: 1330-A Piccard Drive
CITY: Rockville
STATE: Maryland
COUNTRY: U.S.A.
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/876,991
FILING DATE: 16-JUN-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/747,577
FILING DATE:
APPLICATION NUMBER: US/08/650,584
FILING DATE:
APPLICATION NUMBER: US/08/469,702
FILING DATE:
APPLICATION NUMBER: US/08/123,596
FILING DATE:
APPLICATION NUMBER: 07/797,554
FILING DATE: 22-NOV-1991
APPLICATION NUMBER: US 07/494,991
FILING DATE: 16-MAR-1990
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: William M. Blackstone
REGISTRATION NUMBER: 29,772
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 258-5200
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 12284 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Hog cholera virus
STRAIN: Alfort
CELL LINE: PK 15 and 38A1D
FEATURE:
NAME/KEY: CDS
LOCATION: 364..12060
OTHER INFORMATION: /label= 435_kDa_protein
NAME/KEY: primer_bind
LOCATION: complement (2587..2619)
OTHER INFORMATION: /label= primer_1

FEATURE:
NAME/KEY: primer_bind
LOCATION: complement (2842..2880)
OTHER INFORMATION: /label= primer_2
FEATURE:
NAME/KEY: variation
LOCATION: replace(127, "c")
FEATURE:
NAME/KEY: variation
LOCATION: replace(1522, "g")
FEATURE:
NAME/KEY: variation
LOCATION: replace(10989, "t")
US-08-876-991-1

Query Match 56.0%; Score 14; DB 2; Length 12284;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 tctgctgttgaag 23
|||||
Db 2253 TCTGCTGTGGAAG 2266

RESULT 37
US-09-059-853-1
Sequence 1, Application US/09059853
Patent No. 5935582
GENERAL INFORMATION:
APPLICANT: Gregor Meyers, Tillmann R menapf,
APPLICANT: Heinz-J rgen Thiel
TITLE OF INVENTION: Hog cholera virus vaccine and diagnostic
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Organon Teknika Corporation
ADDRESSEE: Biotechnology Research Institute
STREET: 1330-A Piccard Drive
CITY: Rockville
STATE: Maryland
COUNTRY: U.S.A.
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/059,853
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/797,554
FILING DATE: 22-NOV-1991
APPLICATION NUMBER: US 07/494,991
FILING DATE: 16-MAR-1990
ATTORNEY/AGENT INFORMATION:
NAME: William M. Blackstone
REGISTRATION NUMBER: 29,772
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 258-5200
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 12284 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORGANISM: Hog cholera virus
STRAIN: Alfort
CELL LINE: PK 15 and 38A1D

FEATURE:
NAME/KEY: CDS
LOCATION: 364..12060
OTHER INFORMATION: /label= 435_kDa_protein
FEATURE:
NAME/KEY: primer_bind
LOCATION: complement (2587..2619)
OTHER INFORMATION: /label= primer_1
FEATURE:
NAME/KEY: primer_bind
LOCATION: complement (2842..2880)
OTHER INFORMATION: /label= primer_2
FEATURE:
NAME/KEY: variation
LOCATION: replace(127, "c")
FEATURE:
NAME/KEY: variation
LOCATION: replace(1522, "g")
FEATURE:
NAME/KEY: variation
LOCATION: replace(10989, "t")
US-09-059-853-1

Query Match 56.0%; Score 14; DB 2; Length 12284;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 tctgctgttgaag 23
|||||
Db 2253 TCTGCTGTGGAAG 2266

RESULT 38
US-08-592-874-1/c
Sequence 1, Application US/08592874
Patent No. 5854034
GENERAL INFORMATION:
APPLICANT: POLLOCK, THOMAS J.
APPLICANT: YAMAZAKI, MOTOHIDE
APPLICANT: THORNE, LINDA
APPLICANT: MIKOLAJCZAK, MARCIA
APPLICANT: ARMENTROUT, RICHARD W.
TITLE OF INVENTION: DNA SEGMENTS AND METHODS FOR INCRPASING
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: JULES E. GOLDBERG
STREET: 261 MADISON AVENUE
CITY: NEW YORK
STATE: NY
COUNTRY: USA
ZIP: 10016-2391
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/592,874
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/377,440
FILING DATE: 24-JAN-1995
ATTORNEY/AGENT INFORMATION:
NAME: GOLDBERG, JULES E.
REGISTRATION NUMBER: 24,408
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-986-4090
TELEFAX: 212-818-9479
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:

LENGTH: 28804 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
FRAGMENT TYPE: N-terminal
US-08-592-874-1

Query Match
Best Local Similarity 100.0%; Score 14; DB 2; Length 28804;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 tgctggatctgctg 16
|||||
Db 14585 TGCTGGATCTGCTG 14572

RESULT 39
US-09-096-942-2/c
Sequence 2, Application US/09096942
Patent No. 6027925
GENERAL INFORMATION:
APPLICANT: Pollock, Thomas J
APPLICANT: Mikolajczak, Marcia
APPLICANT: Yamazaki, Motohide
APPLICANT: Thorne, Linda
TITLE OF INVENTION: Production of Xanthan Gum by *Sphingomonas Bacteria*
TITLE OF INVENTION: Carrying Genes from *Xanthomonas Campestris*
FILE REFERENCE: seq list for appl filed from pro. appl
CURRENT APPLICATION NUMBER: US/09/096,942
EARLIER FILING DATE: 1998-06-12
EARLIER APPLICATION NUMBER: 60/049,428
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 28804
TYPE: DNA
ORGANISM: *Sphingomonas* sp. S88
US-09-096-942-2

Query Match
Best Local Similarity 100.0%; Score 14; DB 3; Length 28804;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 tgctggatctgctg 16
|||||
Db 14585 TGCTGGATCTGCTG 14572

RESULT 40
US-09-096-867-2/c
Sequence 2, Application US/09096867
Patent No. 6030817
GENERAL INFORMATION:
APPLICANT: Pollock, Thomas J
APPLICANT: Mikolajczak, Marcia
APPLICANT: Yamazaki, Motohide
APPLICANT: Thorne, Linda
TITLE OF INVENTION: Production of Xanthan Gum by *Sphingomonas Bacteria*
TITLE OF INVENTION: Carrying Genes from *Xanthomonas Campestris*
FILE REFERENCE: seq list for appl filed from pro. appl
CURRENT APPLICATION NUMBER: US/09/096,867
EARLIER FILING DATE: 1998-06-11
EARLIER APPLICATION NUMBER: 60/049,428
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2

LENGTH: 28804
TYPE: DNA
ORGANISM: *Sphingomonas* sp. S88
US-09-096-867-2

Query Match
Best Local Similarity 100.0%; Score 14; DB 3; Length 28804;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 tgctggatctgctg 16
|||||
Db 14585 TGCTGGATCTGCTG 14572

RESULT 41
US-08-998-416-727
Sequence 727, Application US/08998416
Patent No. 6239264
GENERAL INFORMATION:
APPLICANT: Philippsen, Peter
APPLICANT: Pohlmann, Rainer
APPLICANT: Steiner, Sabine
APPLICANT: Mohr, Christine
APPLICANT: Wendland, Jürgen
APPLICANT: Knechtle, Philipp
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF *ASHBYA GOSYPII*
TITLE OF INVENTION: AND USES THEREOF
NUMBER OF SEQUENCES: 1152
CORRESPONDENCE ADDRESS:
ADDRESSER: No. 6239264artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: No. 6239264th Carolina
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/998,416
FILING DATE: 24-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 727:
SEQUENCE CHARACTERISTICS:
LENGTH: 749 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: PAG1475RP
US-08-998-416-727

Query Match
Best Local Similarity 100.0%; Score 13; DB 4; Length 749;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 tgctgtttgaagc 24

Db 96 TGCTGTTGAAGC 108

RESULT 42

US-08-153-848-31
Sequence 31, Application US/08153848
Patent No. 5759804

GENERAL INFORMATION:

APPLICANT: Godiska, Ronald

APPLICANT: Gray, Patrick W.

TITLE OF INVENTION: No. 5759804el Seven Transmembrane Receptors

NUMBER OF SEQUENCES: 64

CORRESPONDENCE ADDRESS:

ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &

ADDRESSEE: Bicknell

STREET: 6300 Sears Tower, 233 South Wacker Drive

CITY: Chicago

STATE: Illinois

COUNTRY: USA

ZIP: 60606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/153,848

FILING DATE:

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/977,452

FILING DATE: 17-NOV-1992

ATTORNEY/AGENT INFORMATION:

NAME: No. 5759804and, Greta E.

REGISTRATION NUMBER: 35,302

REFERENCE/DOCKET NUMBER: 31794

TELECOMMUNICATION INFORMATION:

TELEPHONE: (312) 474-6300

TELEFAX: (312) 474-0448

TELEX: 25-3856

INFORMATION FOR SEQ ID NO: 31:

SEQUENCE CHARACTERISTICS:

LENGTH: 1161 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

FEATURE:

NAME/KEY: exon

LOCATION: 7..80

FEATURE:

NAME/KEY: CDS

LOCATION: 94..1158

US-08-153-848-31

Query Match

Best Local Similarity 52.0%; Score 13; DB 1; Length 1161;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 8 gatctgctgttg 20

Db 322 GATCTGCTGTTTG 334

RESULT 43

APPLICANT: Schweikart, Vicki L.

TITLE OF INVENTION: Novel Seven Transmembrane Receptors

NUMBER OF SEQUENCES: 64

CORRESPONDENCE ADDRESS:

ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &

ADDRESSEE: Bicknell

STREET: 6300 Sears Tower, 233 South Wacker Drive

CITY: Chicago

STATE: Illinois

COUNTRY: USA

ZIP: 60606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US93/11153

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/977,452

FILING DATE: 17-NOV-1992

ATTORNEY/AGENT INFORMATION:

NAME: Noland, Greta E.

REGISTRATION NUMBER: 35,302

REFERENCE/DOCKET NUMBER: 31794

TELECOMMUNICATION INFORMATION:

TELEPHONE: (312) 474-6300

TELEFAX: (312) 474-0448

TELEX: 25-3856

INFORMATION FOR SEQ ID NO: 31:

SEQUENCE CHARACTERISTICS:

LENGTH: 1161 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

FEATURE:

NAME/KEY: exon

LOCATION: 7..80

FEATURE:

NAME/KEY: CDS

LOCATION: 94..1158

PCT-US93-11153-31

Query Match

Best Local Similarity 52.0%; Score 13; DB 5; Length 1161;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 8 gatctgctgttg 20

Db 322 GATCTGCTGTTTG 334

RESULT 44

US-08-252-966B-13/c

Sequence 13, Application US/08252966B

Patent No. 5624818

GENERAL INFORMATION:

APPLICANT: Eisenman, Robert N.

APPLICANT: Huth, Peter J.

APPLICANT: Ayer, Donald E.

TITLE OF INVENTION: Regulatory Proteins that Dimerize with

Search completed: October 9, 2001, 15:55:26
 Job time: 13202 sec

```

: ZIP: 98101-2347
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/252,966B
: FILING DATE: 01-JUN-1994
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Shelton, Dennis K.
: REGISTRATION NUMBER: 26,997.
: REFERENCE/DOCKET NUMBER: FHCRI7694
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (206) 682-8100
: TELEFAX: (206) 224-0779
: INFORMATION FOR SEQ ID NO: 13:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1242 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA to mRNA
: DESCRIPTION: clone 10; see Figure 24
: HYPOTHEICAL: NO
: ORIGINAL SOURCE:
: ORGANISM: Mus musculus
: US-08-252-966B-13

```

```

Query Match          52.0%; Score 13; DB 1; Length 1242;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 11 ctgctgtltaag 23
      |||||
Db 525 CTGCTGTTGAG 513

```

```

RESULT 45
US-09-188-930-68/c
: Sequence 68, Application US/09188930A
: Patent No. 6150502
: GENERAL INFORMATION:
: APPLICANT: Watson, James D.
: APPLICANT: Strachan, Lorna
: APPLICANT: Sleeman, Matthew
: APPLICANT: Onrust, Rene
: APPLICANT: Murison, James Greg
: TITLE OF INVENTION: Compositions Isolated From Skin Cells
: FILE REFERENCE: 11000.1011c1
: CURRENT APPLICATION NUMBER: US/09/188,930A
: NUMBER OF SEQ ID NOS: 348
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 68
: LENGTH: 1729
: TYPE: DNA
: ORGANISM: mouse
: US-09-188-930-68

```

```

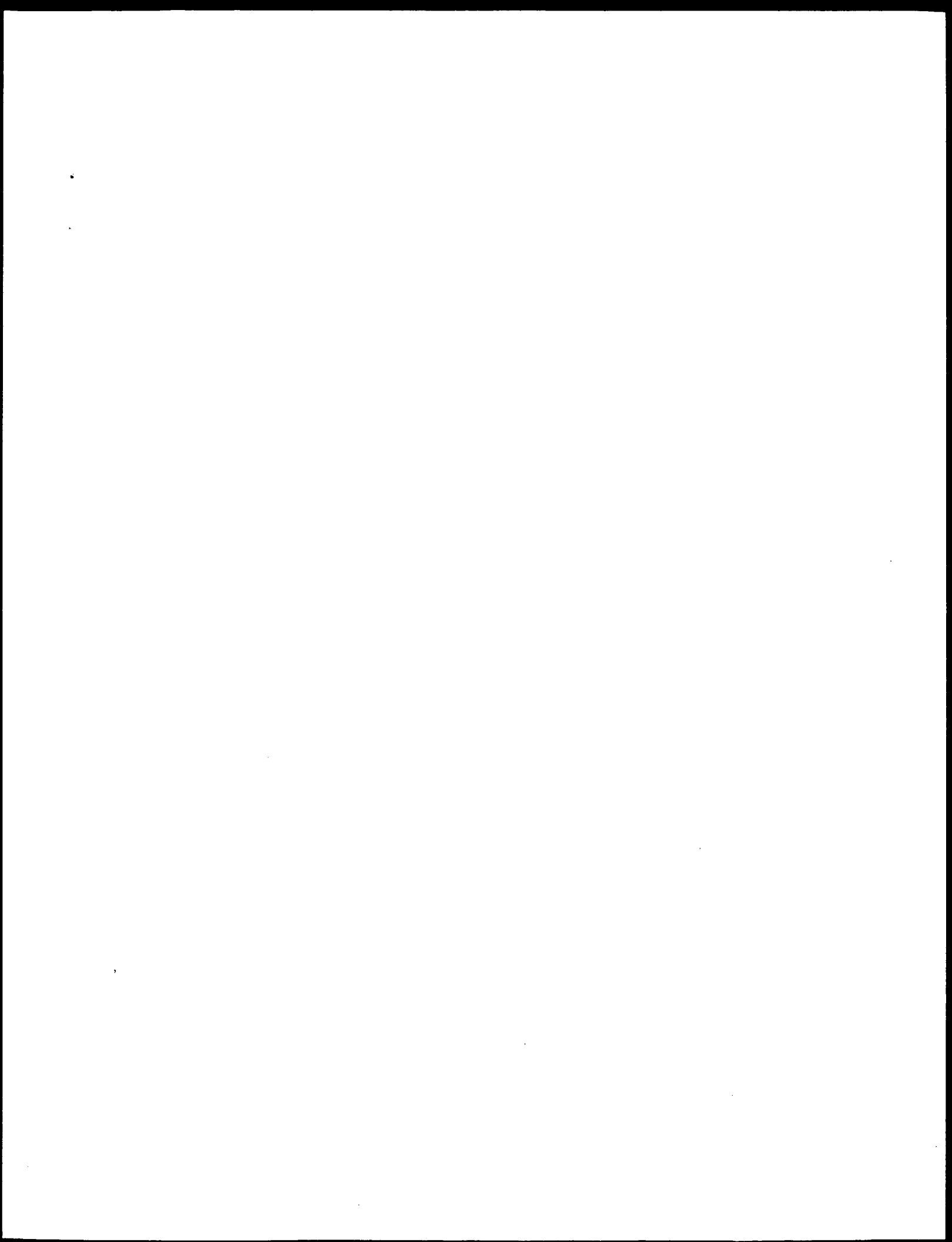
Query Match          52.0%; Score 13; DB 3; Length 1729;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 5 ctgactgtctgt 17
      |||||
Db 1130 CTGACTGCTGT 1118

```

Wed Oct 10 07:45:47 2001

us-09-396-196f-3.oli.rst

Page 1

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 15:15:23 ; Search time 5323.87 Seconds
(without alignments) 44.389 Million cell updates/sec

Title: US-09-396-196f-3

Perfect score: 25
Sequence: 1 gtgtctgagatctgtctgttgaagcg 25

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 10228115 seqs, 4726426750 residues

Word size : 9

Total number of hits satisfying chosen parameters: 754579

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est33:*
26: gb_est34:*
27: gb_est35:*
28: gb_est36:*
29: gb_est37:*
30: gb_est38:*
31: gb_est39:*
32: gb_est40:*
33: gb_est41:*
34: em_estfun:*
35: em_esthum1:*
36: em_esthum2:*
37: em_esthum3:*
38: em_esthum4:*
39: em_esthum5:*
40: em_esthum6:*
41: em_esthum7:*
42: em_esthum8:*
43: em_esthum9:*

44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_esthum29:*
64: em_esthum30:*
65: em_esthum31:*
66: em_esthum32:*
67: em_esthum33:*
68: em_esthum34:*
69: em_esthum35:*
70: em_esthum36:*
71: em_esthum37:*
72: em_esthum38:*
73: em_esthum39:*
74: em_esthum40:*
75: em_esthum41:*
76: em_esthum42:*
77: em_esthum43:*
78: em_esthum44:*
79: em_esthum45:*
80: em_esthum46:*
81: em_esthum47:*
82: em_esthum48:*
83: em_esthum49:*
84: em_esthum50:*
85: em_esthum51:*
86: em_esthum52:*
87: em_esthum53:*
88: em_esthum54:*
89: em_esthum55:*
90: em_esthum56:*
91: em_esthum57:*
92: em_esthum58:*
93: em_esthum59:*
94: em_esthum60:*
95: em_esthum61:*
96: em_esthum62:*
97: em_esthum63:*
98: em_esthum64:*
99: em_esthum65:*
100: em_esthum66:*
101: em_esthum67:*
102: em_esthum68:*
103: em_esthum69:*
104: em_esthum70:*
105: em_esthum71:*
106: em_esthum72:*
107: em_esthum73:*
108: em_esthum74:*
109: em_esthum75:*
110: em_esthum76:*
111: em_esthum77:*
112: em_esthum78:*
113: em_esthum79:*
114: em_esthum80:*
115: em_esthum81:*
116: em_esthum82:*

```
117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*
```

```
190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	19	76.0	888	221	CNS04KH2	AL294815 Tetraodon
2	18	72.0	403	163	BE095496	BE095496 UI-R-BU0
3	18	72.0	487	148	BF428015	BF428015 daa04g08.
4	17	68.0	546	236	A0955256	A0955256 Sheared D
5	17	68.0	547	32	AV1718673	AV1718673 AV1718673
6	16	64.0	237	111	AM101291	AM101291 sq77b02.Y
7	16	64.0	247	162	BE023256	BE023256 sm80a02.Y
8	16	64.0	274	148	BF414722	BF414722 UI-R-BJ2
9	16	64.0	287	134	BB455974	BB455974 BB455974
10	16	64.0	287	171	BF917535	BF917535 113-07011
11	16	64.0	293	240	AZ214122	AZ214122 Sheared D
12	16	64.0	301	114	AM306120	AM306120 se46b06.Y
13	16	64.0	328	138	BE659333	BE659333 GM700009A
14	16	64.0	342	115	AM423279	AM423279 sh65d11.Y
15	16	64.0	360	156	C61070	C61070 C61070 Yuj1
16	16	64.0	365	138	BE662000	BE662000 bs02d05.Y
17	16	64.0	422	24	A1736967	A1736967 sb35c02.Y
18	16	64.0	434	119	AM704839	AM704839 sk40b11.Y
19	16	64.0	436	18	A1293622	A1293622 lp06802.5
20	16	64.0	441	164	BE190258	BE190258 sc09f08.Y
21	16	64.0	470	105	AL373871	AL373871 MBB03D01
22	16	64.0	474	119	AM704996	AM704996 sk41c02.Y
23	16	64.0	481	105	AL379973	AL379973 MBB48E08
24	16	64.0	497	149	BF461669	BF461669 UI-M-CG0P
25	16	64.0	502	136	BE515828	BE515828 WHE0606.B
26	16	64.0	503	105	AL383090	AL383090 MRCB11C02
27	16	64.0	517	105	AL373872	AL373872 MBB03D01
28	16	64.0	517	232	A0692054	A0692054 HS_5408_A
29	16	64.0	521	148	BF419551	BF419551 UI-R-CA0-
30	16	64.0	522	102	AL854720	AL854720 UI-M-BH0-
31	16	64.0	525	105	AL379974	AL379974 MBB48E08
32	16	64.0	531	115	AM395303	AM395303 sh46d09.Y
33	16	64.0	531	115	AM395303	AM395303 HS_2204_A
34	16	64.0	555	235	AM573807	AM573807 EST316398
35	16	64.0	574	148	BE999070	BE999070 EST430793
36	16	64.0	575	150	BF525150	BF525150 UI-R-AF0-
37	16	64.0	580	150	BF525150	BF525150 UI-R-AF0-
38	16	64.0	591	120	AM734894	AM734894 sk76d03.Y
39	16	64.0	601	240	AZ267971	AZ267971 RPCT-23-1
40	16	64.0	609	113	AM267738	AM267738 EST305866
41	16	64.0	616	123	AM980891	AM980891 EST392044
42	16	64.0	621	115	BM585328	BM585328 EST487092
43	16	64.0	621	115	AM422111	AM422111 f157a12.Y
44	16	64.0	627	137	BE593142	BE593142 WSI_98_A0
45	16	64.0	632	244	AZ449344	AZ449344 1M0247704
				16	A1108765	A1108765 GH08003.5

ALIGNMENTS

RESULT 1
CNS04KH2 888 bp DNA GSS 21 MAY-2000
LOCUS Tetraodon nigroviridis genome survey sequence T7 end of clone
DEFINITION 11603 of library G from Tetraodon nigroviridis, genomic survey sequence.

ACCESSION AL294815 GI:8033395
VERSION AL294815.1
KEYWORDS GSS: genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.
REFERENCE 1 (bases 1 to 888)
AUTHORS Roest-Crolius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,

TITLE
JOURNAL
REFERENCE
AUTHORS
Tetraodon nigroviridis DNA sequence

TITLE
JOURNAL
REFERENCE
AUTHORS
Tetraodon nigroviridis DNA sequence

COMMENT
Submitted (12-APR-2000) to the EMBL/Genbank/DBJ databases
this sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the tetraodon nigroviridis
genome. For more information, please take a look at
<http://www.genoscope.cns.fr/tetraodon>.

FEATURES
Source
Location/Qualifiers
1..888
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="11603"
/clone_11b="G"
/note="Genoscope sequence ID : C0BGL16CB02LPI-end : T7"
BASE COUNT 129 a 276 c 276 g 187 t 20 others
ORIGIN

Query Match 76.0%; Score 19; DB 221; Length 888;
Best Local Similarity 100.0%; Pred. No. 1;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 tgcgtgacgcgtgttga 21
|||||
DB 331 TCGTGCATCTGCTGTGA 349

RESULT 2
BE095496/c 403 bp mRNA EST 12-JUN-2000
LOCUS BE095496 UI-R-BU0-apa-c-11-0-UI.s1 UI-R-BU0 Rattus norvegicus cDNA clone
DEFINITION UI-R-BU0-apa-c-11-0-UI 3', mRNA sequence.
ACCESSION BE095496
VERSION BE095496.1 GI:8486427
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 403)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITILE Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu

Oligo-dT track not found. Not 1 site shown in beginning of sequence
is likely internal to the message. cDNA library Preparation: M.B.
Soares Lab Clone distribution: clones will be available through
Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=No.
Location/Qualifiers

```

source
1. .403
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-800-apa-c-11-0-UI"
/clone_lib="UI-R-800"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-800
library is a subtracted library derived from a mixture of
eye and ganglia tissues. For a detailed description of the
library from which this clone was derived, please visit
our web site at ratest.eng.uiowa.edu. The subtraction has
been previously described in (Bonaldi, Lennon and Soares,
Genome Research 6:791-806, 1996)
TAG:560=None found"

BASE COUNT      85 a      147 c      141 g      30 t
ORIGIN

Query Match
Best Local Similarity 100.0%; Score 18; DB 163; Length 403;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtgtcgtgactcgtctgt 18
|||||
Db 245 GTTCTGATCTGCTGTT 228

RESULT 3
BF428015 487 bp mRNA EST 30-MAR-2001
LOCUS daa04q08.x1 NICHD XGC Ovl Xenopus laevis cDNA clone IMAGE:4055247
DEFINITION 3' similar to TR:093478 O93478 KINESIN LIKE PROTEIN 3. ;, mRNA
Sequence.
ACCESSION BF428015 GI:11439676
VERSION BF428015
KEYWORDS EST.
SOURCE African clawed frog.
ORGANISM Xenopus laevis
REFERENCE 1 (bases 1 to 487)
AUTHORS Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D.,
, Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person
, B., Gibbons,M., Harvey,N., Rittler,E., Jackson,Y., McCann,R.,
Waterston,R. and Wilson,R.
Washu Xenopus EST project, 1999
Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Martha Rebert, Steven L. Klein, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: Xenopus clones from this library are available
through the I.M.A.G.E. Consortium/LNL at: info@image.lnl.gov
Seq primer: -40UP from Gibco
High quality sequence stop: 50.
Location/Qualifiers
1. .487
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone="IMAGE:4055247"
/clone_lib="NICHD XGC Ovl"
/sex="female"
/lab_host="DH10B (phage-resistant)"
/note="Organ: ovary; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2.0 kb. Constructed by Life
Technologies."

```

```

BASE COUNT      138 a      101 c      102 g      146 t
ORIGIN

Query Match
Best Local Similarity 100.0%; Score 18; DB 148; Length 487;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtgtcgtgactcgtctgt 18
|||||
Db 249 GTTCTGATCTGCTGTT 266

RESULT 4
AQ952526 546 bp DNA GSS 27-JAN-2000
LOCUS Sheared DNA-48D19,TF Sheared DNA Trypanosoma brucei genomic clone
DEFINITION Sheared DNA-48D19, DNA sequence.
ACCESSION AQ952526
VERSION AQ952526.1 GI:6775791
KEYWORDS GSS.
SOURCE Trypanosoma brucei.
ORGANISM Trypanosoma brucei.
Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma.
REFERENCE 1 (bases 1 to 546)
AUTHORS El-Sayed,N., Zhao,S., Zhao,H., Gall,S., Sub,E., Malek,J., Fujita,C.,
Gerrard,C., Leech,V., de Jong,P., Ullu,E., Melville,S., Donelson,J.,
Fraser,C. and Adams,M.
Determination of clone end sequences from Trypanosoma brucei cDNA
10.1 sheared DNA library
Unpublished (1999)
COMMENT Contact: Najib M. El-Sayed
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: nelsayed@tigr.org
Clones are derived from the Trypanosoma brucei cDNA 10.1 sheared
DNA library constructed at TIGR. Clones will be available for
distribution through ATCC. Sheared DNA end sequences search page:
http://www.tigr.org/tdb/mdb/tbdb/.
Seq primer: M13-forward
Class: Shotgun.
FEATURES
source
1. .546
Location/Qualifiers
/organism="Trypanosoma brucei"
/strain="TREU927/4 cDNA 10.1"
/db_xref="taxon:5691"
/clone="Sheared DNA-48D19"
/clone_lib="Sheared DNA"
/note="Vector: pUC18; Site_1: SmaI; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from a cloned population of
Trypanosoma brucei (TREU927/4 cDNA 10.1) was mechanically
sheared to give a tight size distribution (approx 2 kb).
The v + 1 method used for the library construction is
described in detail in Smith, H.O. and Venter, J.C.
(Making small insert libraries for whole genome shotgun
sequencing projects. In Genome Sequencing: A Practical
Approach, eds. M. Vaudin and B. Barrell, Oxford University
Press, 1999)."
BASE COUNT      192 a      161 c      127 g      66 t
ORIGIN

Query Match
Best Local Similarity 100.0%; Score 17; DB 236; Length 546;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtgtcgtgactcgtctgt 17
|||||

```


Db 192 GTTCTGATCTGCTGT 176

RESULT 5
LOCUS AV718673 547 bp mRNA EST 16-OCT-2000
DEFINITION AV718673 GLC Homo sapiens cDNA clone GLCXYA12 5', mRNA sequence.
ACCESSION AV718673
VERSION AV718673.1 GI:10815825
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 547)
AUTHORS Qian,B., Wu,T., Huang,Q., Kang,B., Gao,X., Xu,Z., Xiao,H., Xu,X., Li,N., Peng,Y., Liu,F., Qu,J., Song,H., Cheng,Z., Zeng,L., Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Yang,Y., Gu,Y., Chen,Z. and Han,Z.
TITLE Homo sapiens cDNA GLC clones
JOURNAL Unpublished (2000)
CONTACT Zequang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhongliang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919 (ex.45)
Fax: 86-21-50801922
Email: hanzq@chgc.sh.cn
This clone is available at CHGC in Shanghai.
FEATURES
source
1..547
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="GLCXYA12"
/clone_lib="GLC"
/tissue_type="corresponding non cancerous liver tissue"
/dev_stage="Adult"
/lab_host="SOLR"
/note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 183 a 101 c 116 g 147 t

ORIGIN

Query Match 68.0%; Score 17; DB 32; Length 547;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 gttctgcatctgctgt 17
|||||

Db 473 GTTCTGATCTGCTGT 457

RESULT 6
LOCUS AM101291 237 bp mRNA EST 06-DEC-1999
DEFINITION sd77b02.y1 Gm-c1009 glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-c1009-388 5', similar to TR:Q40380 ARABINOGLACTIN-PROTEIN PRECURSOR.; mRNA sequence.
ACCESSION AM101291
VERSION AM101291.1 GI:6071904
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.
REFERENCE 1 (bases 1 to 237)
AUTHORS Shoemaker,R., Keim,P., Vodkin,L., Ereding,J., Corryell,V., Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wille,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,M., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann

TITLE
JOURNAL
COMMENT
,R., Waterston,R. and Wilson,R.
Public Soybean EST Project
Unpublished (1999)
Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com
Seq primer: 40RP from Glibco.
FEATURES
source
1..237
Location/Qualifiers
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1009-388"
/clone_lib="Gm-c1009"
/lab_host="XL10-Gold"
/note="Vector: pBluescript II XR; Site_1: EcoRI; Site_2: XhoI; The mRNA was isolated from entire roots of 2-month-old 'Williams' plants that were greenhouse grown in 5-gallon pots. To suppress nodulation, Black Gold All-Purpose potting soil was supplemented with: 0.36g/L available phosphoric acid (P205), 20mg/L urea N, 0.16g/L S, 0.49mg/L B, 2.5mg/L Cu, 0.15g/L Fe, 13.5mg/L Mn, 0.26mg/L Mo, 14mg/L Zn, 20mg/L Ca, and the following nutrients in a slow-release form (Osmocote): 0.165g/L ammonia N, 0.185g/L nitrate N, 0.35g/L available phosphoric acid, and 0.35g/L soluble potash. No nodules were visible on the roots at harvest. Stragene's cDNA Synthesis Kit (catalog #200401) was used to synthesize the cDNA. First-strand synthesis was performed with 5-methyl dCTP, hence the ligated cDNA is hemimethylated. Stragene's first-strand synthesis primer was used (GAGAGAGAGAGAGAGACTGCTGAG(T)-181. After second-strand synthesis is, the cDNA ends were 'polished' with clone Pfu DNA polymerase, ligated to EcoRI adapters, and phosphorylated. The XhoI site within the first-strand synthesis primer was restricted by digestion with XhoI; all XhoI sites in the cDNA would be protected by their hemimethylated status. The cDNA constructs were size-fractionated with a 400bp cutoff, using a SizeSep 400 spun column from Pharmacia. The column eluent was then ligated into Stragene's pBluescript II XR predigested vector (pBluescript II SK(+)) that had been digested with EcoRI and XhoI, and phosphorylated). Both the white and blue colonies appear to contain recombinant plasmids with cDNA inserts. This library was constructed by Dr. Paul Keim and Dr. Virginia Corryell."

BASE COUNT 40 a 60 c 52 g 85 t

ORIGIN

Query Match 64.0%; Score 16; DB 111; Length 237;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 gttctgcatctgctg 16
|||||

Db 96 GTTCTGATCTGCTG 111

RESULT 7
LOCUS BE023256 247 bp mRNA EST 21-NOV-2000
DEFINITION sm80a02.y1 Gm-c1015 glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-c1015-6339 5', mRNA sequence.
ACCESSION BE023256

VERSION BE023256.1 GI:8285697
 KEYWORDS EST
 SOURCE soybean.
 ORGANISM Glycine max
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.
 1 (bases 1 to 247)
 REFERENCE Shoemaker R., Kelm, P., Vodkin, L., Erpelidg, J., Corryell, V., Khanna A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritzer, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.
 Public Soybean EST project
 Unpublished (1999)
 TITLE JOURNAL
 COMMENT Contact: Shoemaker R/Public Soybean EST Project
 Public Soybean EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: estevatson.wustl.edu
 This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com
 Insert Length: 438 Std Error: 0.00
 Seq primer: -40RP from Gibco.
 FEATURES
 source
 Location/Qualifiers
 1..247
 /organism="Glycine max"
 /db_xref="taxon:3847"
 /clone="GENOME SYSTEMS CLONE ID: Gm-cl015-6339"
 /clone_1lb="Gm-cl015"
 /tissue_type="Mature flowers, field grown plants"
 /lab_host="X110-Gold"
 /note="Vector: Bluescript II XR; Site 1: EcoRI; Site 2: XhoI; This cDNA library was constructed from mRNA isolated from mature flowers of field grown plants. The cDNA library was prepared using the Stratagene Bluescript II XR cDNA library construction kit. Complementary DNA was synthesized from mRNA using a primer consisting of a poly (dT) sequence with a XhoI restriction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the Bluescript vector. The ligated cDNA fragments were transformed into X110-Gold host cells. This library was constructed by Dr. Randy Shoemaker and Dr. John Erpelidg."
 BASE COUNT 39 a 65 c 57 g 77 t 9 others
 ORIGIN
 Query Match 64.0%; Score 16; DB 162; Length 247;
 Best Local Similarity 100.0%; Pred. No. 46;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 gtgtgtgagatctgctg 16
 |||
 Db 128 GTTGCTGATCTGCTG 143
 RESULT 8
 BF414722/c 274 bp mRNA EST 28-NOV-2000
 LOCUS BF414722
 DEFINITION UI-R-BJ2-bow-c-11-0-UI.s1 UI-R-BJ2 Rattus norvegicus cDNA clone
 ACCESSION BF414722
 VERSION BF414722.1 GI:11402711
 KEYWORDS EST

SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 1 (bases 1 to 274)
 REFERENCE Bonaldo, M.F., Lennon, G. and Soares, M.B.
 Normalization and subtraction: two approaches to facilitate gene discovery
 Genome Res. 6 (9), 791-806 (1996)
 TITLE JOURNAL
 MEDLINE
 COMMENT Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msoares@blue.weeg.uiowa.edu
 Oligo-dT track not found. Not 1 site shown in beginning of sequence is likely internal to the message. cDNA library Preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com)
 Seq primer: M13 Forward
 POLYA-No.
 FEATURES
 source
 Location/Qualifiers
 1..274
 /organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone="UI-R-BJ2-bow-c-11-0-UI"
 /clone_1lb="UI-R-BJ2"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pUT3D-Pac (Pharmacia) with a modified polylinker. Site 1: Not I; Site 2: Eco RI; The UI-R-BJ2 library is a subtracted library derived from the following tissues: heart, atrium at 15 dpc, ventricle at 16.5 dpc, atrium at 16.5 dpc, ventricle at 13 dpc, ventricle at 15 dpc, AV canal at 15 dpc. For a detailed description of the library from which this clone was derived, please visit our web site at ratest.eng.uiowa.edu. The subtraction has been previously described in (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)
 TAG_SEQ=None found"
 BASE COUNT 54 a 96 c 106 g 18 t
 ORIGIN
 Query Match 64.0%; Score 16; DB 148; Length 274;
 Best Local Similarity 100.0%; Pred. No. 46;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 gtgtgtgagatctgctg 16
 |||
 Db 246 GTTGCTGATCTGCTG 231
 RESULT 9
 BB455974 287 bp mRNA EST 21-JUL-2000
 LOCUS BB455974
 DEFINITION BB455974 RIKEN full-length enriched, 12 days embryo spinal ganglion
 ACCESSION BB455974
 VERSION BB455974.1 GI:9351467
 KEYWORDS EST
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 287)
 REFERENCE Konno, H., Aizawa, K., Akahira, S., Akiyama, J., Arikawa, T., Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N.,

REFERENCE 1 (bases 1 to 293)
AUTHORS El-Sayed, N., Zhao, S., Zhao, H., Gill, S., Suh, E., Malek, J., Fujii, C., Gerrard, C., Leech, V., de Jong, P., Ulliu, E., Melville, S., Donelson, J., Fraser, C., and Adams, M.
TITLE Determination of clone end sequences from Trypanosoma brucei GUTat 10.1 sheared DNA library
JOURNAL Unpublished (1999)
COMMENT Other-GSSs: Sheared DNA-67G1.TR
Contact: Najib M. El-Sayed
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: nelsayed@tigr.org

Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared DNA library constructed at TIGR. Clones will be available for distribution through Research Genetics, Alabama, USA. Sheared DNA end sequences search page: <http://www.tigr.org/tldb/mbd/tldb/>.
Seq primer: M13-Forward
Class: shotgun.

FEATURES
source
1..293
/organism="Trypanosoma brucei"
/strain="TREU927/4 GUTat 10.1"
/db_xref="taxon:5691"
/clone="Sheared DNA-67G1"
/note="Vector: pUC18; Site: 1: SmaI; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from a cloned population of Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared to give a tight size distribution (approx 2 kb). The v + l method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Bartell, Oxford University Press, 1999)."

BASE COUNT
ORIGIN 38 a 73 c 78 g 104 t

Query Match 64.0%; Score 16; DB 240; Length 293;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 atctgcgtctgaagc 24
|||||
Db 149 ATCTGCTGTTGAAGC 164

RESULT 12
AM306120 301 bp mRNA EST 20-JAN-2000
LOCUS
DEFINITION se46b06.y1 Gm-cl017 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
ACCESSION AM306120
VERSION AM306120.1 GI:6718473
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.
1 (bases 1 to 301)
Shoemaker, R., Keim, P., Vodkin, L., Erpelidng, J., Corryell, V., Khanna, A., Bolla, B., Matra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wille, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Rittler, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.

TITLE Public Soybean EST Project
JOURNAL Unpublished (1999)
COMMENT Contact: Shoemaker R./Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu

This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info.genomesystems.com web site: www.genomesystems.com
Seq primer: 40RP from Gibco.

FEATURES
source
1..301
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-cl017-1932"
/clone_lib="Gm-cl017"
/tissue_type="vegetable buds of field grown plants"
/lab_host="XL10-Gold"
/note="Vector: pBluescript II Xr; Site: 1: EcoRI; Site: 2: XhoI; This cDNA library was constructed from mRNA isolated from vegetable buds of field grown plants. The cDNA library was prepared using the Stratagene pBluescript II Xr library construction kit. Complementary DNA was synthesized from mRNA using a primer consisting of a poly (dT) sequence to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into XL10-Gold host cells. This library was constructed by Dr. Randy Shoemaker and Dr. John Erpelidng."

BASE COUNT
ORIGIN 73 a 48 c 60 g 119 t 1 others

Query Match 64.0%; Score 16; DB 114; Length 301;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtctgcgactctcgc 16
|||||
Db 45 GTTGCTGATCTGCTG 60

RESULT 13
BE659333 328 bp mRNA EST 06-SEP-2000
LOCUS
DEFINITION GM700009A20A6 Gm-r1070 Glycine max cDNA clone Gm-r1070-3084 3',
ACCESSION BE659333
VERSION BE659333.1 GI:9985225
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.
1 (bases 1 to 328)
Vodkin, L., Keim, P., Shoemaker, R., Retzel, E., Khanna, A., Corryell, V., Erpelidng, J., Raph, C., Shoop, E., Pardinas, J., Liu, L. and Lewin, H.
A Functional Genomics Program for Soybean (NSF 9872565)
Unpublished (1999)
Other-ESTs: AW423279 corresponding to Gm-cl015-4654 (5')
Contact: Vodkin, L.O., PI. A Functional Genomics Program for Soybean (NSF 9872565)
Lewin, H. A., Director, Keck Center for Comparative and Functional Genomics

University of Illinois
Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA
Tel: (217) 244-6147
Fax: (217) 333-4582
Email: l-vodkin@uiuc.edu

This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134. For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genome systems.com web site: www.genomesystems.com
Seq primer: 5'-TTTTTTTTTTTTTTTTT(A/C/G)-3',

FEATURES

source

Location/Qualifiers

1..328

/organism="Glycine max"

/db_xref="taxon:3847"

/clone="Gm-r1070-3084"

/clone.lib="Gm-r1070"

/note="The library Gm-r1070 is a sequence-driven, retracted set of 9,216 clones selected from cDNA libraries from various tissues and stages of development of soybean that represent 2,639 sequences from immature cotyledons, 1,770 from immature seed coats, 3,938 from flowers, and 869 from young pods. The 5' ESTs of the source clones from the different libraries was used to select singletons, or a representative of each clone, which were retracted to form library Gm-r1070. The cDNA clones of the retracted Gm-r1070 library were then sequenced at the 3' end. The contig analysis to select unique genes was performed by the laboratory of Ernest Reitzel, Center for Computational Genomics and Bioinformatics, University of Minnesota, http://www.cbc.umn.edu/research/projects/soybean/index.html. Reracking was performed by Genome Systems, St. Louis, http://www.genomesystems.com, and 3' sequencing by the Keck Center for Comparative and Functional Genomics, University of Illinois, http://www.life.uiuc.edu/biokech/keck.html. Note: The corresponding 5' EST from each clone in the Gm-r1070 library is listed in the 'OTHER EST' field. The detailed information on the source library for each clone can also be obtained by referring to the Genome Systems clone ID of the original cDNA library that is also listed under 'OTHER EST'."

BASE COUNT 115 a 71 c 76 g 53 t 13 others
ORIGIN

Query Match 64.0%; Score 16; DB 138; Length 328;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gtgtctgagatctgctg 16
|||||
Db 226 GTTCTGATCTGCTG 211

RESULT 14
AM423279 342 bp mRNA EST 09-FEB-2000
LOCUS sh5d11.v1 Gm-cl015 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
DEFINITION ARBINGALACTAN-PROTEIN PRECURSOR. ; mRNA sequence.

ACCESSION AM423279
VERSION AM423279.1 GI:6951280
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max

REFERENCE Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta: Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Rosidae: eurosids I: Fabales: Fabaceae: Papilionoideae: Glycine.
1 (bases 1 to 342)
Shoemaker, R., Keim, P., Vodkin, L., Erpelting, J., Coryell, V., Khanna, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers

TITLE
JOURNAL
COMMENT

'Y', Person, B., Swaller, T., Gibbons, M., Page, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.
Public Soybean EST Project
Unpublished (1999)
Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genome systems.com web site: www.genomesystems.com
Seq primer: -40RP from Gibco
High quality sequence stop: 328.

FEATURES

source

Location/Qualifiers

1..342

/organism="Glycine max"

/db_xref="taxon:3847"

/clone="GENOME SYSTEMS CLONE ID: Gm-cl015-4654"

/clone.lib="Gm-cl015"

/issue.type="Mature flowers, field grown plants"

/lab_host="X110-Gold"

/note="Vector: pBluescript II XE; Site 1: EcoRI; Site 2: XhoI; This cDNA library was constructed from mRNA isolated from mature flowers of field grown plants. The cDNA library was prepared using the Stratagene pBluescript II XE cDNA library construction kit. Complementary DNA was synthesized from mRNA using a primer consisting of a poly (dT) sequence with a XhoI restriction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into X110-Gold host cells. This library was constructed by Dr. Randy Shoemaker and Dr. John Erpelting."

BASE COUNT 70 a 72 c 70 g 130 t
ORIGIN

Query Match 64.0%; Score 16; DB 115; Length 342;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gtgtctgagatctgctg 16
|||||
Db 101 GTTCTGATCTGCTG 116

RESULT 15
C61070/c 360 bp mRNA EST 22-SEP-1997
LOCUS C61070 yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
DEFINITION clone yk224c/ 5', mRNA sequence.

ACCESSION C61070
VERSION C61070.1 GI:2419775
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.

REFERENCE Eukaryota: Metazoa: Nematoda: Chromadorea: Rhabditida; Rhabditioidea ; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 360)
Kohara, Y., Motobash, T., Tabara, H., Watanabe, H., Sugimoto, A., Sano, M., Miyata, A. and Nishigaki, A.
Expression map of the C.elegans genome
Unpublished (1996)
Contact: Yuji Kohara
Gene Library Lab

National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.

FEATURES

source

Location/Qualifiers
1. 360
/organism="Caenorhabditis elegans"
/strain="CB1489 him-8(e1489)"
/db_xref="taxon:6239"
/clone="yk224c7"
/clone_lib="Yuji Kohara unpublished cDNA"
/sex="hermaphrodite, male"
/tissue_type="whole animal"
/dev_stage="varied"
BASE COUNT 101 a 71 c 76 g 111 t 1 others
ORIGIN

Query Match 64.0%; Score 16; DB 156; Length 360;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 atctgcgtttgaagc 24
|||||
Db 26 ATCTGCTGTTGAAGC 11

RESULT 16
BE662000/c 365 bp mRNA EST 08-JAN-2001
DEFINITION bs02d05.y1 Drosophila melanogaster adult testis library Drosophila
melanogaster cDNA clone bs02d05 5', mRNA sequence.
ACCESSION BE662000
VERSION BE662000.1 GI:9992572
KEYWORDS EST.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster

SOURCE

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 365)
Andrews, J., Bouffard, G.G., Cheadle, C., Lu, J., Becker, K.G. and
Oliver, B.

Gene discovery using computational and microarray analysis of
transcription in the drosophila melanogaster testis
Genome Res. 10 (12), 2030-2043 (2000)
20568492
CONTACT: Brian Oliver
Laboratory of Cellular and Developmental Biology
NIDDK, National Institutes of Health
6 Center Drive MSC 2715, Bldg 6, Rm B1-13, Bethesda, MD 20892 USA
Fax: (301) 496 5239
Email: oliver@helix.nih.gov,
http://www.nidk.nih.gov/intram/people/boliver.htm

Tissue isolation and library construction performed at the National
Institute of Diabetes and Digestive and Kidney Diseases, NIH (see
http://www.nidk.nih.gov/intram/people/boliver.htm). DNA sequencing
and analyses performed by National Institutes of Health Intramural
Sequencing Center (NISC; see http://www.nisc.nih.gov).
Plate: 02 row: d column: 05
Seq primer: M13p1 reverse primer (ABI).

FEATURES

source

Location/Qualifiers
1. 365
/organism="Drosophila melanogaster"
/strain="y1[*] w[67c1]/y"
/db_xref="taxon:7227"
/clone="bs02d05"
/clone_lib="Drosophila melanogaster adult testis library"
/sex="male"
/dev_stage="1-5 day adult"
/lab_host="SOLR (Stratagene)"
/note="Organ: testis; Vector: pBluescript SK (Stratagene);

BASE COUNT 85 a 129 c 94 g 56 t 1 others
ORIGIN

Query Match 64.0%; Score 16; DB 138; Length 365;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gttgcgtgacgcgcg 16
|||||
Db 330 GTTGCATGATCTGCTG 315

RESULT 17
A1736967

LOCUS sb35c02.y1 Gm-c1013 glycine max cDNA clone GENOME SYSTEMS CLONE ID:
DEFINITION Gm-c1013-27 5', similar to TR:Q40380 Q40380 ARABINOGLACTAN-PROTEIN
PRECUSOR. ; mRNA sequence.
ACCESSION A1736967
VERSION A1736967.1 GI:5058435
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max

SOURCE

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eustosids I; Fabales; Fabaceae; Papilionoideae; glycine.
1 (bases 1 to 422)

REFERENCE Shoemaker, R., Keim, P., Vodka, L., Erpelting, D., Corvelli, V., Khana
'A', Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C.,
Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers
'Y', Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk
'R', Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann
'R', Waterston, R. and Wilson, R.
Public Soybean EST Project
Unpublished (1999)

CONTACT: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu

This clone is available through: Genome Systems, Inc. 4633 World
Parkway Circle St. Louis, Missouri 63134 For further information
call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)
427-3324 or contact: clones@genomesystems.com or
info@genomesystems.com web site: www.genomesystems.com
Seq primer: -40R from Gibco
High quality sequence stop: 409.

FEATURES

source

Location/Qualifiers
1. 422
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1013-27"
/clone_lib="Gm-c1013"
/tissue_type="Whole seedlings, 2-3 week old seedlings,
greenhouse grown"
/lab_host="X110-Gold"
/note="Vector: pBluescript II XR; Site_1: EcoRI; Site_2:
XhoI; This cDNA library was constructed from mRNA isolated
from whole seedlings of 2-3 week old greenhouse grown

plants. The cDNA library was prepared using the Stratagene pluscript II XR cDNA library construction kit. Complementary DNA was synthesized from mRNA using a primer consisting of a poly (dT) sequence with a XhoI restriction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pluscript vector. The ligated cDNA fragments were transformed into XL10-Gold host cells. This library was constructed by Dr. Randy Shoemaker and Dr. John Erpelidg.

BASE COUNT 83 a 105 c 91 g 138 t 5 others
ORIGIN

Query Match 64.0%; Score 16; DB 24; Length 422;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gttgctgagctcgtc 16
|||||
Db 183 GTTGCTGATCTGCTG 198

RESULT 18
AM704839 434 bp mRNA EST 21-NOV-2000
LOCUS
DEFINITION AK40811.Y1 Gm-c1019 glycine max cDNA clone GENOME SYSTEMS CLONE ID:
AM704839 Gm-c1019-5014 5', mRNA sequence.
ACCESSION
VERSION AM704839.1 GI:7589057
KEYWORDS
SOURCE soybean.
ORGANISM
glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.
1 (bases 1 to 434)

REFERENCE
AUTHORS Shoemaker, R., Kelm, P., Vocklin, L., Erpelidg, J., Coryell, V., Khanna, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Page, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shih, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.
Public Soybean EST Project
Unpublished (1999)
Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810

TITLE
JOURNAL
COMMENT
FEATURES
SOURCE
1. 434
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1019-5014"
/clone_1id="Gm-c1019"
/tissue_type="Immature seed coats of greenhouse grown plants"
/lab_host="DH10B (Gibco BRL)"
/note="Vector: pSPORT1 (Life Technologies); Site_1: Not I; Site_2: Sal I; This cDNA library was constructed from mRNA isolated from immature seed coats (200-300 mgs) of
Email: est@watson.wustl.edu
This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com
Insert Length: 553 Std Error: 0.00
Seq primer: -40RP from Gibco
High quality sequence stop: 413.
Location/Qualifiers

greenhouse grown plants. The library was prepared using the life technologies pluscript II XR cDNA library construction kit. Complementary DNA was synthesized from mRNA using a poly (dT) sequence with a Not I restriction site. Sal I linker adapters were ligated to the blunt-ended cDNA fragments followed by Not I digestion. The cDNA fragments were directionally cloned into the Not I-Sal I restriction site of the pSPORT1 vector. The ligated cDNA fragments were transformed into E.coli ElectroMax DH10B host cells (Gibco BRL). This library was constructed by Dr. Lila Vocklin and Dr. Ann Khanna.

BASE COUNT 57 a 169 c 90 g 118 t
ORIGIN

Query Match 64.0%; Score 16; DB 119; Length 434;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gttgctgagctcgtc 16
|||||
Db 398 GTTGCTGATCTGCTG 413

RESULT 19
A1293622 436 bp mRNA EST 01-DEC-1998
LOCUS
DEFINITION LP06802.5prime LP Drosophila melanogaster larval-early pupal POT2
A1293622 Drosophila melanogaster cDNA clone LP06802 5prime, mRNA sequence.
ACCESSION
VERSION A1293622.1 GI:3943029
KEYWORDS
SOURCE fruit fly.
ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 436)

REFERENCE
AUTHORS Harvey, D., Hong, L., Evans-Holm, M., Pendleton, J., Su, C., Brokstein, R., Lewis, S. and Rubin, G.M.
BDGF/HMT Drosophila EST Project
Unpublished (1997)
Contact: Harvey, D.
G. M. Rubin-Molecular and Cell Biology
University of California Berkeley
539 LSA, Berkeley, CA 94720-3200, USA
Fax: 510 643 9947
Email: http://www.fruitfly.org/EST, estfruitfly.berkeley.edu
Plate: 68 row: A column: 2
High quality sequence stop: 390.
Location/Qualifiers

FEATURES
SOURCE
1. 436
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="LP06802"
/clone_1id="LP Drosophila melanogaster larval-early pupal POT2"
/sex="male and female"
/dev_stage="larvae-pupae"
/lab_host="DH5-alpha"
/note="Organ: whole body; Vector: POT2; Site_1: EcoRI; Site_2: XhoI; Sized fractionated cDNAs were directly ligated into POT2. Plasmid cDNA library.
Location/Qualifiers

BASE COUNT 111 a 122 c 111 g 92 t
ORIGIN

Query Match 64.0%; Score 16; DB 18; Length 436;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gttgctgagctcgtc 16

Db 328 GTTGCTGATCTGCTG 313

RESULT 20

LOCUS

BE190258

DEFINITION BE190258 441 bp mRNA EST 22-JUN-2000

ACCESION Gm-c1035-2848 5', mRNA sequence.

VERSION BE190258

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

LOCATION/Qualifiers

source

1. 441

/organism="Glycine max"

/db_xref="taxon:3847"

/clone="GENOME SYSTEMS CLONE ID: Gm-c1035-2848"

/clone_lib="Gm-c1035"

/tissue_type="immature leaves of greenhouse grown plants"

/lab_host="DH10B"

/note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; This

CDNA library was constructed from mRNA isolated from

immature leaves (unfurling trifoliolate) of greenhouse grown

plants that were 2 weeks old. The library was prepared

using the Life Technologies Superscript CDNA library

construction kit. Complementary DNA was synthesized from

mRNA using a poly (dT) sequence with a NotI restriction

site. SalI linker adapters were ligated to the

blunt-ended CDNA fragments followed by NotI digestion.

The CDNA fragments were directionally cloned into the

NotI-SalI restriction site of the pSPORT1 vector. The

ligated CDNA fragments were transformed into E.coli

Electromax DH10B host cells. This library was constructed

in the laboratory of Dr. Lila Vodkin by Anu Khanna at the

University of Illinois at Urbana-Champaign. e-mail:

l-vodkin@uiuc.edu"

BASE COUNT

ORIGIN

102 a

63 c

85 g

190 t

1 others

Query Match

Best Local Similarity

Matches

16; Conservative

Db 48 GTTGCTGATCTGCTG 63

RESULT 21

LOCUS

AL373871

DEFINITION AL373871 470 bp mRNA EST 03-AUG-2000

ACCESION MTRB03D01F1 MTRB Medicago truncatula cDNA clone MTRB03D01 T3, mRNA

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

LOCATION/Qualifiers

source

1. 470

/organism="Medicago truncatula"

/cultivar="Jemalong"

/db_xref="taxon:3880"

/clone="MTRB03D01"

/clone_lib="MTRB"

/tissue_type="Symbiotic root nodules"

/dev_stage="harvested 4 days post inoculation with

Sinorhizobium meliloti"

/note="Vector: pBluescript PSK; Site_1: EcoRI; Site_2:

XhoI; Plants were grown in an aeroponic chamber on

nitrogen-rich medium for 21 days. Three days before

inoculation with Sinorhizobium meliloti, the medium was

replaced by N-free medium. Root nodules (+ short adjacent

root segments) were harvested 4 days post inoculation.

CDNA was prepared from polyA+ enriched RNA. The CDNA was

directionally ligated into Uni-zap XR vector from

stratagene and packaged using Gigapack Gold packaging

extracts. Plasmids containing CDNA inserts were

mass-excised from phage stocks using Exsacit helper phage

and propagated in S01R cells. Clone ordering and

sequencing was performed by the Centre National de

Sequencage (Genoscope, Evry, France)."

BASE COUNT

ORIGIN

85 a

137 c

68 g

180 t

Query Match

Best Local Similarity

Matches

16; Conservative

0; Mismatches

0; Indels

0; Gaps

0;

RESULT 22

LOCUS

AM704996

DEFINITION AM704996 474 bp mRNA EST 18-APR-2000

SK41c02.y1 Gm-c1019 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:

ACCESSION AM704996
 VERSION AM704996.1 GI:7589217
 KEYWORDS EST
 SOURCE soybean.
 ORGANISM Glycine max

REFERENCE Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta: Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Rosidae: eurosids I: Fabales: Fabaceae: Papilionoideae: Glycine. 1 (bases 1 to 474)

AUTHORS Shoemaker, R., Keim, P., Vodkin, L., Erpelding, J., Corryell, V., Khanna, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Rittler, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.

TITLE Public Soybean EST Project

JOURNAL Unpublished (1999)

COMMENT Contact: Shoemaker R./Public Soybean EST Project
 Public Soybean EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.wustl.edu
 This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com
 Seq primer: -40RP from Gibco
 High quality sequence stop: 434.

FEATURES
 source
 Location/Qualifiers
 1..474
 /organism="Glycine max"
 /db_xref="taxon:3847"
 /clone="GENOME SYSTEMS CLONE ID: GM-cl019-5115"
 /clone_1lb="GM-cl019"
 /tissue_type="Immature seed coats of greenhouse grown plants"
 /lab_host="DH10B (Gibco BRL)"
 /note="Vector: pSPORT1 (Life Technologies); Site_1: Not I; Site_2: Sal I; This cDNA library was constructed from mRNA isolated from immature seed coats (200-300 mgs) of greenhouse grown plants. The library was prepared using the Life Technologies superscript cDNA library construction kit. Complementary DNA was synthesized from mRNA using a poly (dT) sequence with a Not I restriction site. Sal I linker adapters were ligated to the blunt-ended cDNA fragments followed by Not I digestion. The cDNA fragments were directionally cloned into the Not I-Sal I restriction site of the pSPORT1 vector. The ligated cDNA fragments were transformed into E.coli ElectroMax DH10B host cells (Gibco BRL). This library was constructed by Dr. Jila Vodkin and Dr. Anu Khanna."

BASE COUNT
 66 a 175 c 104 g 129 t

ORIGIN

Query Match 64.0%; Score 16; DB 119; Length 474;
 Best Local Similarity 100.0%; Pred. No. 48;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gtgtctgactgtctg 16
 |||
 DB 398 GTTGCTGATCTGCTG 413

RESULT 23
 LOCUS AL379973 481 bp mRNA EST 03-AUG-2000
 DEFINITION MCB48B08F1 MCB Medicago truncatula cDNA clone MCB48B08 T3, mRNA sequence.

ACCESSION AL379973
 VERSION AL379973.1 GI:9679725
 KEYWORDS EST
 SOURCE barrel medic.
 ORGANISM Medicago truncatula

REFERENCE Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta: Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Rosidae: eurosids I: Fabales: Fabaceae: Papilionoideae: Medicago. 1 (bases 1 to 481)

AUTHORS Journeel, E.P., Crespeau, H., van Tuinen, D., Gouzy, J., Tallon, O., Niebel, A., Carreau, V., Chataigner, O., Kahn, D., Gianinazzi-Pearson, V. and Gamas, P.

TITLE Medicago truncatula ESTs from Sinorhizobium meliloti-induced root nodules

JOURNAL Unpublished (2000)

COMMENT Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: seque@genoscope.cns.fr, Web : www.genoscope.cns.fr
 Contact : Pascal Gamas and Etienne-Pascal Journeel, Laboratoire de Biologie Molculaire des Relations Plantes-Microorganismes, CNRS-INRA, BP 27 31326 Castanet-Tolosan Cedex, France (Email : mt-est@toulouse.inra.fr Website : http://sequence.toulouse.inra.fr/Mtruncatula.html).
 Location/Qualifiers
 1..481
 /organism="Medicago truncatula"
 /cultivar="Jemalong"
 /db_xref="taxon:3880"
 /clone="MTB48B08"
 /clone_1lb="MTBB"
 /tissue_type="symbiotic root nodules"
 /dev_stage="harvested 4 days post inoculation with Sinorhizobium meliloti"
 /note="Vector: pBluescript PSK, Site_1: EcoRI, Site_2: XhoI; Plants were grown in an aeroponic chamber on nitrogen-rich medium for 21 days. Three days before inoculation with Sinorhizobium meliloti, the medium was replaced by N-free medium. Root nodules (+ short adjacent root segments) were harvested 4 days post inoculation. cDNA was prepared from polyA+ enriched RNA. The cDNA was directionally ligated into uni-zap XR vector from Stratagene and packaged using GigaPack Gold packaging extracts. Plasmids containing cDNA inserts were mass-excised from phage stocks using ExAssit helper phage and propagated in SOLR cells. Clone ordering and sequencing was performed by the Centre National de Sequencage (Genoscope, Evry, France)."

BASE COUNT
 89 a 152 c 64 g 176 t

ORIGIN

Query Match 64.0%; Score 16; DB 105; Length 481;
 Best Local Similarity 100.0%; Pred. No. 48;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gtgtctgactgtctg 16
 |||
 DB 385 GTTGCTGATCTGCTG 400

RESULT 24
 LOCUS BF461669/c 497 bp mRNA EST 04-DEC-2000
 DEFINITION UT-M-CGDP-bmu-g-12-0-UI-s1 NRH BMAP Ret4_S2 Mus musculus cDNA clone
 UT-M-CGDP-bmu-g-12-0-UI-3', mRNA sequence.
 ACCESSION BF461669
 VERSION BF461669.1 GI:11530852
 KEYWORDS EST
 SOURCE house mouse.
 ORGANISM Mus musculus

DEFINITION Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi: Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.

REFERENCE 1 (bases 1 to 497)
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 COMMENT Contact: Chlo. H
 National Institute of Mental Health
 6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
 20892-9643, USA
 Tel: 301 443 1706
 Fax: 301 443 9890
 Email: mestr@mail.nih.gov

Oligo-dT track not found, Not 1 site shown in beginning of sequence is likely internal to the message. cDNA Library Preparation: M.B. Soares Lab Clone distribution: Researchers may obtain BMAP cDNA clones from RESEARCH GENERICS. It should be noted that Bento Soares is generating a small number of additional specialized non-redundant arrays of BMAP cDNAs whose availability will be considered under appropriate and limited collaborative arrangements
 Seq primer: M13 Forward
 POLYA-No.

FEATURES
 source Location/Qualifiers

1..497
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="U1-M-CG0P-Dmu-9-12-0-UT"
 /clone_id="NIH_BMAP_Ret4_S2"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pT713D-Pac (Pharmacia) with a modified polylinker. Site 1: Not I; Site 2: Eco RI; The NIH BMAP Ret4_S2 library is a subtracted library ultimately derived from mouse retina tissue libraries at various stages of development. For a detailed description of the library from which this clone was derived, please visit our web site at brainest.eng.uiowa.edu.
 TAG-Seq=None found"
 BASE COUNT 112 a 168 c 171 g 44 t 2 others
 ORIGIN

Query Match 64.0%; Score 16; DB 149; Length 497;
 Best Local Similarity 100.0%; Pred. No. 49;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtctcgtgctgctg 16
 |||||||||||||
 Db 246 GTTCTGATCTGCTG 231

RESULT 25
 BE515828 502 bp mRNA EST 08-AUG-2000
 LOCUS WHE0606.B07.C14ZA Wheat ABA-treated embryo cDNA library Triticum
 DEFINITION aestivum cDNA clone WHE0606.B07.C14, mRNA sequence.
 ACCESSION BE515828
 VERSION BE515828.1 GI:9739842
 KEYWORDS EST.
 SOURCE bread wheat.
 ORGANISM Triticum aestivum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
 1 (bases 1 to 502)
 REFERENCE 1
 Anderson,O.D., Chao,S., Han,P.S., Hsia,C.C., Johnson,R.R., Kang,Y.,
 Lazo,G.R., Miller,R., Rausch,C.J., Seaton,C.L., Tong,J.C., Verhey,
 S.D. and Walker-Simmons,M.K.
 The structure and function of the expressed portion of the wheat
 genomes - ABA-treated embryo library
 Unpublished (2000)
 COMMENT Contact: Glin Anderson
 US Department of Agriculture, Agriculture Research Service, Pacific

West Area, Western Regional Research Center
 800 Buchanan Street, Albany, CA 94710, USA
 Tel: 5105595773
 Fax: 5105595818
 Email: oanderson@pw.usda.gov

Sequence have been trimmed to remove vector sequence and low
 quality sequence with phred score less than 20
 Seq primer: Clontech Matchmaker 3' AD primer.
 Location/Qualifiers

FEATURES
 source

1..502
 /organism="Triticum aestivum"
 /cultivar="Brevor (soft, white, winter, common wheat)"
 /db_xref="taxon:4565"
 /clone="WHE0606.B07.C14"
 /clone_id="Wheat ABA-treated embryo cDNA library"
 /tissue_type="Seed embryo"
 /dev_stage="Mature dormant seeds"
 /lab_host="E. coli DH12S"
 /note="Vector: pGAD10; Site 1: EcoRI; Site 2: XhoI;
 Embryos were cut from mature, dormant seeds and imbedded in
 25 microm ABA (abscisic acid) in 5 mM Mes buffer, pH 5.7,
 for 12 hr at 22 C. The tissue, total RNA, and poly(A)
 RNA were prepared by Steven Verhey in M.K.
 Walker-Simmons's lab (USDA ARS, Washington State Univ.,
 Pullman, Washington 99164-6420. A cDNA library was made
 by Clontech using a combination of random and oligo dT
 primers. Library was plated and archived by Russell
 Johnson (Colby College, ME/Walker-Simmons' lab). Plasmid
 DNA preparations and DNA sequencing were performed in the
 OD Anderson lab (all other authors)."
 BASE COUNT 103 a 108 c 128 g 162 t 1 others
 ORIGIN

Query Match 64.0%; Score 16; DB 136; Length 502;
 Best Local Similarity 100.0%; Pred. No. 49;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 atctgctgttgaagc 24
 |||||||||||||
 Db 189 ATCTGCTGTGTGAAGC 204

RESULT 26
 AL383090 503 bp mRNA EST 03-AUG-2000
 LOCUS MTCB11G02R1 MTCB Medicago truncatula cDNA clone MTCB11G02 T7, mRNA
 DEFINITION sequence.
 ACCESSION AL383090
 VERSION AL383090.1 GI:9682841
 KEYWORDS EST.
 SOURCE batrel medic.
 ORGANISM Medicago truncatula
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Medicago.
 1 (bases 1 to 503)
 REFERENCE 1
 Journet,E.P., Crespeau,H., van Tuinen,D., Gouzy,J., Jaillon,O.,
 Niebel,A., Carreau,V., Chataigner,O., Kahn,D., Gianinazzi-Pearson
 V. and Gamas,P.
 Medicago truncatula ESTs from endomycorrhizal roots
 Unpublished (2000)
 COMMENT Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 Evry cedex - France
 Email: seque@genoscope.cns.fr, Web : www.genoscope.cns.fr
 Contact : Pascal Gamas and Etienne-Pascal Journet, Laboratoire de
 Biologie Moleculaire des Relations Plantes-Microorganismes,
 CNRS-INRA, BP 27 31326 Castanet-Tolosan Cedex, France (Email :
 Mc-est@toulouse.inra.fr Website :
 http://sequence.toulouse.inra.fr/Mtruncatula.html).

FEATURES
 source

1..503
 Location/Qualifiers

/organism="Medicago truncatula"
 /cultivar="Jemalong"
 /db_xref="taxon:3880"
 /clone="MTBC11C02"
 /clone_lib="MTBC"
 /tissue_type="arbuscular mycorrhiza"
 /dev_stage="harvested 3 weeks post inoculation with Glomus intraradices"
 /note="Vector: pBluescript pSK; Site-1: EcoRI; Site-2: XhoI; M. truncatula sterilised seeds were germinated for 72h at 25 C, before transplanting into a 1/3 Epsilones soil : 2/3 calcined Terragreen mix in the presence of onion root fragments colonized by the arbuscular mycorrhizal fungus Glomus intraradices (Schenck & Smith, isolate IPR8). The plants were watered every day and twice a week with a modified nutrient long Ashton solution without phosphate but with a high level of nitrate. After 3 weeks RNA was extracted from whole root systems. cDNA was prepared from polyA+ enriched RNA. The cDNA was directionally ligated into Uni-zap XR vector from Stratagene and packaged using Gigapack Gold packaging extracts. Plasmids containing cDNA inserts were mass-excised from phage stocks using Exasit helper phage and propagated in SOLR cells. Clone ordering and sequencing was performed by the Centre National de Sequencage (Genoscope, Evry, France). Note : EST may be of fungal origin."

BASE COUNT 107 a 139 c 76 g 181 t
 ORIGIN

Query Match 64.0%; Score 16; DB 105; Length 503;
 Best Local Similarity 100.0%; Pred. No. 49;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gtgtcgtgagtcgtcgt 16
 |||||||
 Db 237 GTTGCTGATCTGCTG 252

RESULT 27
 AL373872 517 bp mRNA EST 03-AUG-2000
 LOCUS MEBB03D01.R1 MTBB Medicago truncatula cDNA clone MEBB03D01 T7, mRNA
 DEFINITION
 ACCESSION AL373872
 VERSION AL373872.1 GI:9673624
 KEYWORDS EST.
 SOURCE barrel medic.
 ORGANISM Medicago truncatula
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Medicago. 1 (bases 1 to 517)
 REFERENCE Journal,E.P., Crespeau,H., van Tuinen,D., Gouzy,D., Jalllon,O., Nibbel,A., Carreau,V., Chatagnier,O., Kahn,D., Gianlinazzi-Pearson V. and Gamas,P.
 TITLE Medicago truncatula ESTs from Sinorhizobium meliloti-induced root nodules
 JOURNAL Unpublished (2000)
 COMMENT Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 Evry cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
 Contact : Pascal Gamas and Etienne-Pascal Journeil, Laboratoire de Biologie Moleculaire des Relations Plantes-Microorganismes, CNRS-INRA, BP 27 31326 Castanet-Tolosan Cedex, France (Email : Mt-est@toulouse.inra.fr Website : http://sequence.toulouse.inra.fr/Mtruncatula.html).
 FEATURES
 Source 1..517
 /organism="Medicago truncatula"
 /cultivar="Jemalong"
 /db_xref="taxon:3880"

/clone="MTBB03D01"
 /clone_lib="MTBB"
 /tissue_type="symbiotic root nodules"
 /dev_stage="harvested 4 days post inoculation with Sinorhizobium meliloti"
 /note="Vector: pBluescript pSK; Site-1: EcoRI; Site-2: XhoI; Plants were grown in an aeroponic chamber on nitrogen-rich medium for 21 days. Three days before inoculation with Sinorhizobium meliloti, the medium was replaced by N-free medium. Root nodules (+ short adjacent root segments) were harvested 4 days post inoculation. cDNA was prepared from polyA+ enriched RNA. The cDNA was directionally ligated into Uni-zap XR vector from Stratagene and packaged using Gigapack Gold packaging extracts. Plasmids containing cDNA inserts were mass-excised from phage stocks using Exasit helper phage and propagated in SOLR cells. Clone ordering and sequencing was performed by the Centre National de Sequencage (Genoscope, Evry, France)."

BASE COUNT 104 a 147 c 71 g 195 t
 ORIGIN

Query Match 64.0%; Score 16; DB 105; Length 517;
 Best Local Similarity 100.0%; Pred. No. 49;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gtgtcgtgagtcgtcgt 16
 |||||||
 Db 274 GTTGCTGATCTGCTG 289

RESULT 28
 A0692054 517 bp DNA GSS 06-JUL-1999
 LOCUS HS_5408_A2_E07-T7A RPCI-11 Human Male BAC Library Homo sapiens
 DEFINITION genomic clone Plate=984 Col=14 Row=I, DNA sequence.
 ACCESSION A0692054
 VERSION A0692054.1 GI:5382302
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo. 1 (bases 1 to 517)
 REFERENCE Mahairas,G.C., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,T., Young,J., Zhao,S., Adams,M.D. and Hood,L.
 TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
 JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 MEDLINE 99380589
 COMMENT Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pliet de Jong (pliet@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm) or from Resear h Genetics (info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu
 Plate: 964 Row: I Column: 14
 Seq primer: T7
 Class: BAC ends
 High quality sequence stop: 517.
 FEATURES
 Source 1..517
 Location/Qualifiers
 1..517
 /organism="Homo sapiens"
 /cultivar="Jemalong"
 /db_xref="taxon:9606"

```

/clone="Plate-984 Col=14 Row=1"
/clone_lib="RPCI-11 Human Male BAC library"
/sex="male"
/Note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI Methylase. Size selected DNA was cloned into the
pBAC3.6 vector at EcoRI sites"
BASE COUNT      154 a      96 c      83 g      178 t      6 others
ORIGIN

Query Match      64.0%; Score 16; DB 232; Length 517;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      9 atctgctgttgaagc 24
        |||
Db      411 ATCTGCTGTTCAGAC 426

RESULT 29
BF419551/c      521 bp      mRNA      EST      28-NOV-2000
LOCUS      UI-R-CA0-bpd-c-11-0-UI.s1 UI-R-CA0 Rattus norvegicus cDNA clone
DEFINITION      BF419551
VERSION      BF419551.1 GI:11407540
KEYWORDS      EST.
SOURCE      Norway rat.
ORGANISM      Rattus norvegicus;
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 521)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
JOURNAL      Contact: Soares, MB
MEDLINE      Program for Rat Gene Discovery and Mapping
COMMENT      University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
Oligo-dT track not found. Not 1 site shown in beginning of sequence
is likely internal to the message. cDNA library preparation: M.B.
Soares Lab Clone distribution: clones will be available through
Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA-No.

FEATURES
source
Location/Qualifiers
1..521
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone_lib="UI-R-CA0-bpd-c-11-0-UI"
/lab_host="UI-R-CA0"
/Note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-CA0
library is a subtracted library derived from the following
tissues: thalamus, cerebellum, hypothalamus, medulla, pons
and hippocampus. For a detailed description of the library
from which this clone was derived, please visit our web
site at ratseq.eng.uiowa.edu. The subtraction has been
previously described in (Bonaldo, Lennon and Soares,
Genome Research 6:791-806, 1996)
TAG-Seq=None found"
BASE COUNT      113 a      179 c      183 g      46 t

```

```

ORIGIN

Query Match      64.0%; Score 16; DB 148; Length 521;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 gttgcgcgactctg 16
        |||
Db      246 GTTGCTGATCTGCTG 231

```

```

RESULT 30
A1854720/c      522 bp      mRNA      EST      15-JUL-1999
LOCUS      UI-M-BH0-akc-f-12-0-UI.s1 NIH_BMAP_M.S1 Mus musculus cDNA clone
DEFINITION      UI-M-BH0-akc-f-12-0-UI 3', mRNA sequence.
VERSION      A1854720
KEYWORDS      A1854720.1 GI:5498626
SOURCE      EST.
ORGANISM      house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 522)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
JOURNAL      Contact: Chin, H
MEDLINE      National Institute of Mental Health
COMMENT      6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mwest@nih.nih.gov
Oligo-dT track not found. Not 1 site shown in beginning of sequence
is likely internal to the message. cDNA library preparation: M.B.
Soares Lab Clone distribution: NIH BMAP cDNA clones will be made
available by the means that is soon to be determined. When NIH
determines the means for distribution of the BMAP cDNA clones, this
record will be updated accordingly when that means is determined.
Seq primer: M13 Forward
POLYA-No.

FEATURES
source
Location/Qualifiers
1..522
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_lib="UI-M-BH0-akc-f-12-0-UI"
/lab_host="NIH_BMAP_M.S1"
/dev_stage="27-32 days"
/Note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The
NIH_BMAP_M.S1 library is a subtracted library derived from
a mixture of normalized libraries from ten regions of the
mouse brain (cerebellum, brain stems, olfactory bulbs,
hypothalamus, cortex, amygdala, basal ganglia, pineal
gland, striatum, hippocampus). The diver used for
subtraction consisted of a pool of 20,000 cDNA clones
obtained from non-normalized and normalized libraries of
these ten regions of the mouse brain.
TAG-LIB=NIH_BMAP_M.S1
TAG-TISSUE=hypothalamus
TAG-Seq=CGTA"
BASE COUNT      118 a      174 c      179 g      51 t

```

```

Query Match      64.0%; Score 16; DB 102; Length 522;
Best Local Similarity 100.0%; Pred. No. 49;

```

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gttcgtgagatcgtcgt 16
|||||

Db 245 GTTGCTGATCTGCTG 230

RESULT 31

AL379974 525 bp mRNA EST 03-AUG-2000
LOCUS MEBB48E08R1 MEBB Medicago truncatula cDNA clone MEBB48E08 T7, mRNA
DEFINITION sequence.
ACCESSION AL379974
VERSION AL379974.1 GI:9679726
KEYWORDS EST.
SOURCE barrel medic.
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Medicago.
1 (bases 1 to 525)
REFERENCE 1
Journet,E.P., Crespeau,H., Van-Tuinen,D., Gouzy,J., Jallion,O.,
Niebel,A., Carreau,V., Chatagnier,O., Kahn,D., Gianinazzi-Pearson
V. and Gamas P.
Medicago truncatula ESTs from Sinorhizobium meliloti-induced root
nodules
Unpublished (2000)
Contact: Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Contact : Pascal Gamas and Etienne-Pascal Journet, Laboratoire de
Biologie Moléculaire des Relations Plantes-Microorganismes,
CNRS-INRA, BP 27 31326 Castanet-Tolosan Cedex, France (Email :
ME-est@toulouse.inra.fr Website :
http://sequence.toulouse.inra.fr/Medicago.html).
Location/Qualifiers
1. 525
/organism="Medicago truncatula"
/cultivar="Jemalong"
/db_xref="taxon:3880"
/clone_id="MEBB48E08"
/clone_1lb="MEBB"
/tissue_type="symbiotic root nodules"
/dev_stage="harvested 4 days post inoculation with
Sinorhizobium meliloti"
/note="Vector: pBluescript psk; Site_1: EcoRI; Site_2:
XhoI; Plants were grown in an aeroponic chamber on
nitrogen-rich medium for 21 days. Three days before
inoculation with Sinorhizobium meliloti, the medium was
replaced by N-free medium. Root nodules (+ short adjacent
root segments) were harvested 4 days post inoculation.
cDNA was prepared from polyA+ enriched RNA. The cDNA was
directionally ligated into Uni-zap XR vector from
Stratagene and packaged using Gigapack Gold packaging
extracts. Plasmids containing cDNA inserts were
mass-excised from phage stocks using Exasit helper phage
and propagated in SOLR cells. Clone ordering and
sequencing was performed by the Centre National de
Sequencage (Genoscope, Evry, France)."

BASE COUNT 108 a 148 c 81 g 188 t
ORIGIN

Query Match 64.0%; Score 16; DB 105; Length 525;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gttcgtgagatcgtcgt 16
|||||

Db 255 GTTGCTGATCTGCTG 270

RESULT 32

AM395303 531 bp mRNA EST 17-JUL-2000
LOCUS SH46D09.Y1 Gm-c1017 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
DEFINITION Gm-c1017-4746 5' similar to SW:RFB9.DROME P36558 DNA-DIRECTED RNA
POLYMERASE II 15.1 KD POLYPEPTIDE ;, mRNA sequence.
ACCESSION AM395303
VERSION AM395303
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.
1 (bases 1 to 531)
REFERENCE 1
Shoemaker,R., Keim,P., Vodkin,L., Erpelting,J., Coryell,V., Khanna
A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
Wyllie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
R., Waterston,R. and Wilson,R.
Public Soybean EST Project
Unpublished (1999)
Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
This clone is available through: Genome Systems, Inc. 4633 World
Parkway Circle St. Louis, Missouri 63134 For further information
call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)
427-3334 or contact: clones@genomesystems.com or
info@genomesystems.com web site: www.genomesystems.com
Insert length: 743 Std Error: 0.00
Seq primer: -40RP from Gibco
High quality sequence stop: 342.
Location/Qualifiers
1. 531
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1017-4746"
/clone_1lb="Gm-c1017"
/tissue_type="Vegetable buds of field grown plants"
/lab_host="XL10-Gold"
/note="Vector: pBluescript II XR; Site_1: EcoRI; Site_2:
XhoI; This cDNA library was constructed from mRNA isolated
from vegetable buds of field grown plants. The cDNA
library was prepared using the Stratagene pBluescript II
XR library construction kit. Complementary DNA was
synthesized from mRNA using a primer consisting of a poly
(dT) sequence with a XhoI restriction site. EcoRI adapters
were ligated to the blunt-ended cDNA fragments followed by
XhoI digestion. The cDNA fragments were directionally
cloned into the EcoRI-XhoI restriction site of the
pBluescript vector. The ligated cDNA fragments were
transformed into XL10-Gold host cells. This library was
constructed by Dr. Randy Shoemaker and Dr. John
Erpelting."

BASE COUNT 159 a 96 c 107 g 169 t
ORIGIN

Query Match 64.0%; Score 16; DB 115; Length 531;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gttcgtgagatcgtcgt 16
|||||

Db 60 GTTGCTGATCTGCTG 75

RESULT 33
 LOCUS A0889100/c
 DEFINITION HS.2204_A1-C07_MR_CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=2204 Col=13 Row=E, DNA sequence.
 ACCESSION A0889100
 VERSION A0889100.1 GI:6345290
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 555)
 Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D., and Hood,L.
 TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
 JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 MEDLINE 99380589
 COMMENT Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Clones may be purchased from Research Genetics (info@resgen.com).
 BAC end Web Server: http://www.htsc.washington.edu
 Plate: 2204 row: E column: 13
 Seq primer: M13 Reverse
 Class: BAC ends
 High quality sequence stop: 555.
 FEATURES
 source
 Location/Qualifiers
 1..555
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="CIT Approved Human Genomic Sperm Library D"
 /sex="male"
 /note="Organ: Sperm; Vector: pBelorAC11; BAC Clones in E-Coli DH10B"
 BASE COUNT 131 a 135 c 147 g 124 t 18 others
 ORIGIN
 Query Match 64.0%; Score 16; DB 235; Length 555;
 Best Local Similarity 100.0%; Pred. No. 49;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 10 tctctgttgaacg 25
 |||||||||||||
 Db 497 TCTGCTCTTGACG 482

RESULT 34
 LOCUS AW573807
 DEFINITION EST316398 GVN Medicago truncatula cDNA clone pgVN-48022, mRNA sequence.
 ACCESSION AW573807
 VERSION AW573807.1 GI:7238540
 KEYWORDS EST.
 SOURCE barrel medic.
 ORGANISM Medicago truncatula
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Medicago.
 1 (bases 1 to 574)
 Fedorova,M., Pierson,B.L., Samac,D.A., Vance,C.P., Gantt,J.S., Peng,H., Ellis,L., Town,C.D., Bowman,C.L., Craven,M.B., Hansen,T.S., Holt,I.E., and Fraser,C.M.
 TITLE ESTs from one month old nitrogen-fixing root nodules of Medicago

JOURNAL
 COMMENT unpublished (2000)
 Contact: Carroll P. Vance
 Department of Agronomy and Plant Genetics
 University of Minnesota
 411 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108 USA
 Tel: 612 625 5715
 Fax: 651-649-5058
 Email: Vance004@maroon.tc.umn.edu
 Minnesota EST name: M252705e
 TIGR sequence name: MTCAP957K
 More information is available at: .
 'http://chrystie.tamu.edu/medicago/'
 Seq primer: SKmod (CTA gAA gTg gAT CC).
 FEATURES
 source
 Location/Qualifiers
 1..574
 /organism="Medicago truncatula"
 /cultivar="genotype A17"
 /db_xref="taxon:3880"
 /clone="pgVN-48022"
 /clone_lib="GVN"
 /tissue_type="N2-fixing root nodules"
 /dev_stage="effective root nodules harvested one month post inoculation with Sinorhizobium meliloti"
 /lab_host="E. coli strain XL0LR"
 /note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2: XhoI; cDNA was prepared from polyA+ enriched RNA from effective root nodules harvested one month post inoculation with Sinorhizobium meliloti. The cDNA was directionally ligated into the uni-ZAP Xk vector from stragene and packaged using Gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-ZAP phage using Ex-Assist helper phage and propagated in XL0LR cells."
 BASE COUNT 110 a 165 c 80 g 218 t 1 others
 ORIGIN
 Query Match 64.0%; Score 16; DB 118; Length 574;
 Best Local Similarity 100.0%; Pred. No. 49;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 gttctggaatctgctg 16
 |||||||||||||
 Db 370 GTTCTGATCTGCTG 385

RESULT 35
 LOCUS BE999070
 DEFINITION EST430793 GVSN Medicago truncatula cDNA clone pgVSN-15C19, mRNA sequence.
 ACCESSION BE999070
 VERSION BE999070.1 GI:10699346
 KEYWORDS EST.
 SOURCE barrel medic.
 ORGANISM Medicago truncatula
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Medicago.
 1 (bases 1 to 575)
 Fedorova,M., Pierson,B.L., Samac,D.A., Gantt,J.S., Vance,C.P., Town,C.D., Bowman,C.L., Craven,M.B., Cho,J., and Fraser,C.M.
 TITLE ESTs from senescent nodules of Medicago truncatula
 JOURNAL Unpublished (2000)
 COMMENT Contact: Carroll P. Vance
 Department of Agronomy and Plant Genetics
 University of Minnesota
 411 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108 USA
 Tel: 612 625 5715
 Fax: 651-649-5058
 Email: Vance004@maroon.tc.umn.edu
 University of Minnesota name: M273048e TIGR sequence name:

MTKAY22WK More information is available at:
<http://chryslie.tamu.edu/medicago>
 Seq primer: Skmod (CTA gAA CTA gTg gAT CC).
 Location/Qualifiers

FEATURES

1.575
 /organism="Medicago truncatula"
 /cultivar="genotype A17"
 /db_xref="taxon:3880"
 /clone_lib="pGVSN-15C19"
 /clone_lib="GVSN"
 /tissue_type="senescent root nodules"
 /dev_stage="mixture of effective nodules from 40 day old plants harvested 36 hours post shoot removal and nodules collected from 2 month old plants at mid-pod stage"
 /lab_host="E. coli strain SOLR"
 /note="Vector: pBluescript SK +/-; Site_1: EcoRI; Site_2: XhoI; cDNA was prepared from polyA+ enriched RNA from the mixture of effective nodules of 40 day old plants harvested 36 hours post shoot removal and nodules collected from 2 month old plants at mid-pod stage. The cDNA was directionally ligated into the Uni-ZAP XR vector from Stratagene and packaged using GigaPack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-ZAP phage using Ex-Assist helper phage and propagated in SOLR cells."
 BASE COUNT 103 a 164 c 82 g 226 t
 ORIGIN

Query Match 64.0%; Score 16; DB 143; Length 575;
 Best Local Similarity 100.0%; Pred. No. 49;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gttgtgagatcgtcgt 16
 |||||
 Db 347 GTTGCTGATCTGCTG 362

RESULT 36
 LOCUS BF525150 580 bp mRNA EST 11-DEC-2000
 DEFINITION BF525150-580-bp mRNA from Rattus norvegicus cDNA clone
 UI-R-AF0-yd-e-05-0-UI-1 UI-R-AF0-yd-e-05-0-UI-5', mRNA sequence.
 ACCESSION BF525150
 VERSION BF525150.1 GI:11633117
 KEYWORDS EST.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 580)
 Bernaldo,M.F., Lennon,G. and Soares,M.B.
 Normalization and subtraction: two approaches to facilitate gene
 discovery
 Genome Res. 6 (9), 791-806 (1996)
 97044477
 Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msoares@blue.weeg.uiowa.edu
 cDNA library preparation: M.B. Soares Lab Clone distribution:
 clones will be available through Research Genetics (www.resgen.com)
 This clone is also available through the I.M.A.G.E. Consortium at
 LBNL (info@image.lbnl.gov). IMAGE ID= 1795307
 Seq primer: M13 Forward.

FEATURES
 source 1.580
 /organism="Rattus norvegicus"
 /strain="Sprague-Dawley"

/db_xref="taxon:10116"
 /clone="UI-R-AF0-yd-e-05-0-UI"
 /clone_lib="UI-R-AF0"
 /dev_stage="adult"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pRT3D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-AF0
 library is a non-normalized library constructed from 15
 dpc rat atriocentricular (AV) canal. The tag is a string
 of 5 nucleotides present between the Not I site and the
 oligo-dT track. The library was constructed as described
 by Bernaldo, Lennon and Soares, Genome Research 6: 791-806
 1996. Tissue provided by Jim Lin, Department of Biology,
 University of Iowa."
 BASE COUNT 155 a 146 c 147 g 132 t
 ORIGIN

Query Match 64.0%; Score 16; DB 150; Length 580;
 Best Local Similarity 100.0%; Pred. No. 49;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 tctgtgagatcgtcgt 18
 |||||
 Db 321 TCCTGATCTGCTGCTG 336

RESULT 37
 LOCUS AW734894 591 bp mRNA EST 24-APR-2000
 DEFINITION SK76d03.Y1 Gm-c1016 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
 Gm-c1016-10014 5', mRNA sequence.
 ACCESSION AW734894
 VERSION AW734894.1 GI:7640531
 KEYWORDS EST.
 SOURCE soybean.
 ORGANISM Glycine max
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eustosids I; Fabales; Fabaceae; Papilionoideae; Glycine.

REFERENCE 1 (bases 1 to 591)
 AUTHORS Shoemaker,R., Kelm,P., Vodkin,L., Erpelting,J., Corryell,V., Khanna
 A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
 Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
 Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
 R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
 R., Waterston,R. and Wilson,R.
 Public Soybean EST Project
 Unpublished (1999)
 TITLE Public Soybean EST Project
 JOURNAL Contact: Shoemaker R/Public Soybean EST Project
 COMMENT Public Soybean EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available through: Genome Systems, Inc. 4633 World
 Parkway Circle St. Louis, Missouri 63134 For further information
 call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)
 427-3324 or contact: clones@genomesystems.com or
 info@genomesystems.com web site: www.genomesystems.com
 Seq primer: -40RP from Gibco
 High quality sequence stop: 380.

FEATURES

source 1.591
 /organism="Glycine max"
 /db_xref="taxon:3847"
 /clone="GENOME SYSTEMS CLONE ID: Gm-c1016-10014"
 /clone_lib="Gm-c1016"
 /tissue_type="Immature flowers of field grown plants"
 /lab_host="XL10-Gold"
 /note="Vector: pBluescript II XR; Site_1: EcoRI; Site_2:
 XhoI; This cDNA library was constructed from mRNA isolated

from immature flowers of field grown plants. The cDNA library was prepared using the Stratagene Bluescript II XR library construction kit. Complementary DNA was synthesized from mRNA using a primer consisting of a poly (dT) sequence with a XhoI restriction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the Bluescript vector. The ligated cDNA fragments were transformed into XL10-Gold host cells. This library was constructed by Dr. Randy Shoemaker and Dr. John Erpelting."

BASE COUNT 121 a 180 c 103 g 185 t 2 others

Query Match 64.0%; Score 16; DB 120; Length 591;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ttgtctgattctgtctg 16
|||||

Db 330 GTTGCTGATCTGCTG 345

RESULT 38
A2267971 601 bp DNA GSS 26-JUL-2000
LOCUS RPT-23-103L11.T3 RPT-23 Mus musculus genomic clone RPT-23-103L11
DEFINITION , DNA sequence.
ACCESSION A2267971
VERSION A2267971.1 GI:9481584
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 601)
Zhao, S., Nieman, W., Feldblum, T., Malek, J., Shatsman, S., Aklnret, B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P. and Fraser, C.M.
Mouse BAC End Sequences from Library RPT-23
Unpublished (1999)
Other_GSSs: RPT-23-103L11.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPT-23. For BAC library availability, please contact Pieter de Jong
(pieter@tigr.org, med.bufileo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.bufileo.edu/orderingframe.htm>) or from Resea ch Genetics (<http://resgen.com>). BAC end page: http://www.tigr.org/tdb/bac-ends/mouse/bac_end_intro.html
Plate: 103 row: L column: 11
Seq primer: SP6
Class: BAC ends.

FEATURES
Source Location/Qualifiers
1..601
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPT-23-103L11"
/clone_lib="RPT-23"
/sex="Female"
/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBAC3.6; Site: 1; EcoRI; Site: 2; EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size

selected DNA was cloned into the pBAC3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 189 a 92 c 111 g 209 t

Query Match 64.0%; Score 16; DB 240; Length 601;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 ttgtctgattctgtctg 17
|||||

Db 354 TTGCTGATCTGCTG 369

RESULT 39
AM267738 609 bp mRNA EST 07-SEP-2000
LOCUS EST305866 DSIR Medicago truncatula cDNA clone pDSIR-717, mRNA
DEFINITION sequence.
ACCESSION AM267738
VERSION AM267738.1 GI:6654694
KEYWORDS EST.
SOURCE barrel medic.
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Medicago.
1 (bases 1 to 609)
Fedorova, M., Pierson, B.L., Sanae, D.A., Vance, C.P., Gantt, G.S., Peng, H., Ellis, L., Town, C.D., Bowman, C.L., Craven, M.B., Hansen, T.S., Holt, L.E. and Fraser, C.M.
ESTs from roots of Medicago truncatula after inoculation with Phytophthora medicaginis
Unpublished (1999)
Contact: Carroll P. Vance
Department of Agronomy and Plant Genetics
University of Minnesota
411 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108 USA
Tel: 612 625 5715
Fax: 651-649-5058
Email: vance004@maroon.tc.umn.edu
Minnesota EST name: M250222E
TIGR sequence name: MTBA527K
More information, including clone ordering, is available at: <http://chrysis.tamu.edu/medicago>
Seq primer: SKmod (CTA gAA gTg gAT CC).
Location/Qualifiers
1..609
/organism="Medicago truncatula"
/cultivar="genotype A17"
/db_xref="taxon:3880"
/clone="pDSIR-717"
/clone_lib="DSIR"
/tissue_type="roots infected with Phytophthora medicaginis"
/dev_stage="roots harvested at 10 days post inoculation with Phytophthora medicaginis"
/lab_host="E. coli strain XL0LR"
/note="Vector: pBluescript SK-; Site: 1; EcoRI; Site: 2; XhoI; cDNA was prepared from polyA+ enriched RNA from roots harvested at 10 days post inoculation with Phytophthora medicaginis. The cDNA was directionally ligated into the Uni-ZAP XR vector from Stratagene and packaged using Gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-ZAP phage using EX-Assist helper phage and propagated in XL0LR cells. Note: EST may be of fungal origin."

BASE COUNT 117 a 172 c 82 g 238 t

Query Match 64.0%; Score 16; DB 113; Length 609;
 Best Local Similarity 100.0%; Pred. No. 49;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtctgctgactctgctg 16
 |||
 Db 381 GTTGCTGATCTGCTG 396

RESULT 40
 AM980891 616 bp mRNA EST 07-SEP-2000
 LOCUS EST392004 GVN Medicago truncatula cDNA clone pGVN-60C11, mRNA
 DEFINITION
 SEQUENCE.
 AM980891
 VERSION AM980891.1 GI:8172436
 KEYWORDS
 SOURCE EST.
 ORGANISM
 Medicago truncatula
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Medicago.
 1 (bases 1 to 616)
 REFERENCE
 Fedorova, M., Pierson, B.L., Samac, D.A., Vance, C.P., Gant, J.S., Peng
 H., Ellis, L., Town, C.D., Bowman, C.L., Craven, M.B., Hansen, T.S.,
 Holt, I.E. and Fraser, C.M.
 ESTs from one month old nitrogen-fixing root nodules of Medicago
 truncatula
 TITLE
 Unpublished (2000)
 JOURNAL
 Contact: Carroll P. Vance
 Department of Agronomy and Plant Genetics
 University of Minnesota
 411 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108 USA
 Tel: 612 625 5715
 Fax: 651-649-5058
 Email: Vance004@maroon.tc.umn.edu
 University of Minnesota name: M256725e
 TIGR sequence name: MTCBM18TK
 More information is available at: <http://chrysis.tamu.edu/medicago>
 Seq primer: SKmod (CTA gAA CTA gTg gAT CC).
 FEATURES
 source
 1. 616
 /organism="Medicago truncatula"
 /cultivar="genotype A17"
 /db_xref="taxon:3880"
 /clone="pGVN-60C11"
 /clone_1ib="GVN"
 /tissue_type="N2-fixing root nodules"
 /dev_stage="effective root nodules harvested one month
 post inoculation with Sinorhizobium meliloti"
 /lab_host="E. coli strain XL0LR"
 /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
 XhoI; cDNA was prepared from polyA+ enriched RNA from
 effective root nodules harvested one month post
 inoculation with Sinorhizobium meliloti. The cDNA was
 directionally ligated into the Uni-ZAP XR vector from
 StrataGene and packaged using Gigapack III Gold packaging
 extracts. Plasmids containing cDNA inserts were excised
 from the recombinant lambda-ZAP phage using Ex-Assist
 helper phage and propagated in XL0LR cells."
 BASE COUNT 118 a 173 c 82 g 243 t
 ORIGIN
 Query Match 64.0%; Score 16; DB 123; Length 616;
 Best Local Similarity 100.0%; Pred. No. 50;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtctgctgactctgctg 16
 |||
 Db 388 GTTGCTGATCTGCTG 403

RESULT 41
 BG585328 618 bp mRNA EST 11-APR-2001
 LOCUS EST487092 MHAM Medicago truncatula/Glomus versiforme mixed EST
 DEFINITION
 BG585328
 accession BG585328
 version BG585328.1 GI:13600392
 keywords
 source EST.
 organism Medicago truncatula/Glomus versiforme mixed EST library.
 Medicago truncatula/Glomus versiforme mixed EST library
 Eukaryota; mixed EST libraries.
 1 (bases 1 to 618)
 REFERENCE
 Harrison, M.J., Liu, J., Town, C.D., Van Aken, S., Uterback, T., Cho, J.
 and Fraser, C.M.
 ESTs from roots of Medicago truncatula after colonization with
 Glomus versiforme, 2001
 TITLE
 Unpublished (2001)
 JOURNAL
 Contact: Harrison M.J.
 Plant Biology Division
 The Samuel Roberts Noble Foundation
 2510 Sam Noble Parkway, Ardmore, OK 73401
 Tel: 580-223-5810
 Fax: 580-221-7380
 Email: mjharrison@noble.org
 Noble EST name: N380802e TIGR sequence name: MNDCC35TK More
 information is available at: <http://www.medicago.org>
 Seq primer: SKmod (CTA gAA CTA gTg gAT CC).
 FEATURES
 source
 1. 618
 /organism="Medicago truncatula/Glomus versiforme mixed EST
 library"
 /cultivar="Medicago truncatula genotype A17"
 /db_xref="taxon:119092"
 /clone="PMHAM-22F21"
 /clone_1ib="MHAM"
 /tissue_type="roots colonized with Glomus versiforme"
 /dev_stage="Roots harvested at 10, 17, 22, 31 and 38 days
 post-inoculation with Glomus versiforme. The library was
 made from a mixture of RNA from each of these stages."
 /lab_host="E. coli strain XL0LR"
 /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
 XhoI; cDNA was prepared from polyA+ enriched RNA from
 roots harvested at 10, 17, 22, 31 and 38 days
 post-inoculation with Glomus versiforme. The cDNA was
 directionally ligated into the UniZap XR vector from
 StrataGene and packaged using Gigapack III Gold packaging
 extracts. Plasmids containing cDNA inserts were excised
 from the recombinant lambda-Zap phage using Ex-assist
 helper phage and propagated in XL0LR cells."
 BASE COUNT 119 a 174 c 84 g 241 t
 ORIGIN
 Query Match 64.0%; Score 16; DB 155; Length 618;
 Best Local Similarity 100.0%; Pred. No. 50;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gtctgctgactctgctg 16
 |||
 Db 390 GTTGCTGATCTGCTG 405

RESULT 42
 AM422111 621 bp mRNA EST 09-FEB-2000
 LOCUS fi57a12 y1 Suano Kawakami zebrafish DRB Danio rerio cDNA clone
 DEFINITION 2641726 y1 similar to gb:M99436 TRANSLOCIN-LIKE ENHANCER PROTEIN 2
 (HUMAN);, mRNA sequence.
 AM422111
 VERSION AM422111.1 GI:6950043
 KEYWORDS
 SOURCE EST.
 zebrafish.

ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes; Cyprinidae; Rasbora; Danio.
1 (bases 1 to 621)

REFERENCE Sugano, S., Kawakami, K., Johnson, S., Li, F., Marra, M., Eddy, S., Hillier, L., Clifton, S., Allen, M., Gibbons, M., Jost, S., Kucaba, T., Martin, J., Pape, D., Stepien, K., Underwood, K., Theising, B., Ritter, E., Bowers, Y., Wylie, T., Waterston, R. and Wilson, R.
Masfou Zebrafish EST Project 1999
Unpublished (1999)

TITLE Washington University School of Medicine
JOURNAL 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
COMMENT Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
Library constructed by Dr. Sumio Sugano and Dr. Koichi Kawakami DNA Sequencing by: Washington University Genome Sequencing Center
Seq primer: J3 ET from Amersham
High quality sequence stop: 504.

FEATURES
source Location/Qualifiers
1..621
/organism="Danio rerio"
/strain="AB"
/db_xref="taxon:7955"
/clone="2641726"
/clone_lib="Sugano Kawakami Zebrafish DNA"
/sex="mixed (one male and one female, including unfertilized eggs)"
/dev_stage="adult"
/lab_host="DH10B (phage resistant)"
/note="Vector: pME18S-FL3; Site 1: DraIII (CACTGCTG); Site 2: DraIII (CACCATG); 1st strand cDNA was primed with an oligo(dT) primer [ATGCGCCCTTTTCTTTT]; double-stranded cDNA was ligated to a DraIII adaptor [TGTGGCCCTGCG], digested and cloned into distinct DraIII sites of the pME18S-FL3 vector (5' site CACTGCTG, 3' site CACCATG). XhoI should be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5kb. Library constructed by Dr. Sumio Sugano (University of Tokyo Institute of Medical Science) and kindly donated by Dr. Koichi Kawakami. Custom primers for sequencing: 5' and primer CTTCTGCTCTAAAGCTCG and 3' end primer CGACCTGCAGCTCAGCACA."

BASE COUNT 185 a 174 c 143 g 119 t
ORIGIN

Query Match 64.0%; Score 16; DB 115; Length 621;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 ttgtcgatctgctgt 17
|||||
Db 138 TTGCTGATCTGCTGT 123

RESULT 43
BE593142 627 bp mRNA EST 18-AUG-2000
LOCUS WS1_98_A09.g1_A002 Water-stressed 1 (WS1) Sorghum bicolor cDNA,
DEFINITION mRNA sequence.
ACCESSION BE593142
VERSION BE593142.1 GI:9848215
KEYWORDS EST.
SOURCE Sorghum.
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoidae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 627)
Cordonnier-Pratt, M.-M., Gingle, A., Marsala, C., Sudman, M. and Pratt

TITLE An EST database from Sorghum: water-stressed plants
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmp@atuga.edu
Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.
Seq primer: PolyTMix
High quality sequence start: 55
High quality sequence stop: 627
POLYA-No.

FEATURES
source Location/Qualifiers
1..627
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Water-stressed 1 (WS1)"
/note="Organ: Mix of 5-week old plants on days 7 & 8 after water was withheld; Vector: Lambda Zap; Site 1: XhoI; Site 2: EcoRI; The library was made from poly-A RNA in the cloning vector Lambda Zap II. Clones to be sequenced were prepared by mass excision."

BASE COUNT 169 a 152 c 124 g 182 t
ORIGIN

Query Match 64.0%; Score 16; DB 137; Length 627;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 ttgtcgatctgctgt 18
|||||
Db 208 TGCTGATCTGCTGT 193

RESULT 44
AZ449344 632 bp DNA GSS 04-OCT-2000
LOCUS 1M0247J04 Mouse 10kb plasmid UUGC1M library Mus musculus genomic
DEFINITION clone UUGC1M0247J04 R, DNA sequence.
ACCESSION AZ449344
VERSION AZ449344.1 GI:10603038
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 632)
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tinney, A., von Niederhausen, A. and Wright, D., Weiss, R.
Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0247 row: J column: 04
Seq primer: CACACGGAACACGATATAC
Class: plasmid ends
High quality sequence stop: 632.

FEATURES

source

Location/Qualifiers

1. 632

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="U081024704"

/clone_lib="Mouse 10kb plasmid U081024704"

/sex="Male"

/lab_host="E. coli strain XL10-Gold, T1-resistant, F-"

/note="Vector: PMD29; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PMD42 (g11473211419b/AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT

185 a 167 c 110 g 170 t

ORIGIN

Query Match 64.0%; Score 16; DB 244; Length 632;

Best Local Similarity 100.0%; Pred.No. 50;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ggaatcgtcgtgttga 22

|||||

Db 552 GGATCTGCTGTTCGA 567

RESULT 45

A1108765/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

fruit fly.

Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 640)

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (1997)

Contact: Harvey, D.

G. M. Rubin-Molecular and Cell Biology

University of California Berkeley

539 USA, Berkeley, CA 94720-3200, USA

Fax: 510 643 9947

Email: http://www.fruitfly.org/EST, estfruitfly.berkeley.edu

Plate: 80 row: A column: 3

High quality sequence stop: 412.

FEATURES

source

1. 640

/organism="Drosophila melanogaster"

/db_xref="taxon:7227"

/clone="GH08003"

/clone_lib="GH Drosophila melanogaster head port2"

/sex="male and female"

/dev_stage="adult"

/lab_host="DH5 - alpha"

/note="Organ: head; Vector: port2; Site 1: EcoRI; Site 2: XhoI; Sized fractionated cDNAs were directly ligated into port2 plasmid cDNA library."

BASE COUNT 160 a 211 c 175 g 94 t

ORIGIN

Query Match 64.0%; Score 16; DB 16; Length 640;

Best Local Similarity 100.0%; Pred.No. 50;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gttgctggaatcgtcgtg 16

|||||

Db 352 GTTGCTGATCTGCTG 337

Search completed: October 9, 2001, 15:15:26

Job time: 13648 sec

Wed Oct 10 07:45:47 2001

us-09-396-196f-3.011.rst

Page 24

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 15:44:56 ; Search time 1670.83 Seconds
(without alignments)
231.438 Million cell updates/sec

Title: US-09-396-196F-4

Perfect score: 25
Sequence: 1 ggaatcgtctgtgaagcgacgacg 25

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1344157 seqs, 7733874588 residues

Word size : 9

Total number of hits satisfying chosen parameters: 237205

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl: *
1: gb_ba1: *
2: gb_ba2: *
3: gb_ba3: *
4: gb_in1: *
5: gb_in2: *
6: gb_in3: *
7: gb_om: *
8: gb_ov: *
9: gb_pat1: *
10: gb_pat2: *
11: gb_ph: *
12: gb_pl1: *
13: gb_pl2: *
14: gb_pl3: *
15: gb_pl4: *
16: em_ba1: *
17: em_ba2: *
18: em_fun: *
19: em_htgo_hum: *
20: em_htgo_inv: *
21: em_htgo_rod: *
22: em_htg_hum1: *
23: em_htg_hum2: *
24: em_htg_hum3: *
25: em_htg_hum4: *
26: em_htg_hum5: *
27: em_htg_hum6: *
28: em_htg_hum7: *
29: em_htg_hum8: *
30: em_htg_inv1: *
31: em_htg_inv2: *
32: em_htg_other: *
33: em_htg_rod: *
34: em_hum1: *
35: em_hum2: *
36: em_hum3: *
37: em_hum4: *
38: em_hum5: *
39: em_hum6: *
40: em_hum7: *
41: em_in: *
42: em_om: *
43: em_or: *

44: em_ov: *
45: em_pat: *
46: em_ph: *
47: em_pl: *
48: em_ro: *
49: em_sts: *
50: em_sy: *
51: em_un: *
52: em_v1: *
53: gb_sts1: *
54: gb_sts2: *
55: gb_sts3: *
56: gb_sy: *
57: gb_un: *
58: gb_v1: *
59: gb_v12: *
60: gb_hcg1: *
61: gb_hcg2: *
62: gb_hcg3: *
63: gb_hcg4: *
64: gb_hcg5: *
65: gb_hcg6: *
66: gb_hcg7: *
67: gb_hcg8: *
68: gb_hcg9: *
69: gb_hcg10: *
70: gb_hcg11: *
71: gb_hcg12: *
72: gb_hcg13: *
73: gb_hcg14: *
74: gb_hcg15: *
75: gb_hcg16: *
76: gb_hcg17: *
77: gb_hcg18: *
78: gb_hcg19: *
79: gb_hcg20: *
80: gb_hcg21: *
81: gb_hcg22: *
82: gb_hcg23: *
83: gb_hcg24: *
84: gb_hcg25: *
85: gb_pr1: *
86: gb_pr2: *
87: gb_pr3: *
88: gb_pr4: *
89: gb_pr5: *
90: gb_pr6: *
91: gb_pr7: *
92: gb_pr8: *
93: gb_pr9: *
94: gb_pr1: *
95: gb_pr2: *
96: gb_pr4: *
97: gb_pr10: *
98: em_ba3: *

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	25	100.0	1041	9	AR029499	AR029499 Sequence
2	25	100.0	1041	9	AR034916	AR034916 Sequence
3	25	100.0	1084	9	A11530	A11530 B108 gene O
4	25	100.0	5526	2	AF250776	AF250776 Uncultured
5	25	100.0	5793	2	ECOBIO	J04423 E.coli 7,8-
6	25	100.0	5872	9	A38246	A38246 Sequence 1
7	25	100.0	5872	9	A38251	A38251 Sequence 6
8	25	100.0	5872	9	A93674	A93674 Sequence 1

```

9      25      5872      9      A93679 Sequence 6
10     25      5872      9      AR101809 Sequence
11     25      5872      9      AR101810 Sequence
12     25      5872      9      AR101810 Sequence
13     25      5872      9      AR101810 Sequence
14     25      5872      9      AR101810 Sequence
15     25      5872      9      AR101810 Sequence
16     25      5872      9      AR101810 Sequence
17     25      5872      9      AR101810 Sequence
18     25      5872      9      AR101810 Sequence
19     25      5872      9      AR101810 Sequence
20     25      5872      9      AR101810 Sequence
21     25      5872      9      AR101810 Sequence
22     25      5872      9      AR101810 Sequence
23     25      5872      9      AR101810 Sequence
24     25      5872      9      AR101810 Sequence
25     25      5872      9      AR101810 Sequence
26     25      5872      9      AR101810 Sequence
27     25      5872      9      AR101810 Sequence
28     25      5872      9      AR101810 Sequence
29     25      5872      9      AR101810 Sequence
30     25      5872      9      AR101810 Sequence
31     25      5872      9      AR101810 Sequence
32     25      5872      9      AR101810 Sequence
33     25      5872      9      AR101810 Sequence
34     25      5872      9      AR101810 Sequence
35     25      5872      9      AR101810 Sequence
36     25      5872      9      AR101810 Sequence
37     25      5872      9      AR101810 Sequence
38     25      5872      9      AR101810 Sequence
39     25      5872      9      AR101810 Sequence
40     25      5872      9      AR101810 Sequence
41     25      5872      9      AR101810 Sequence
42     25      5872      9      AR101810 Sequence
43     25      5872      9      AR101810 Sequence
44     25      5872      9      AR101810 Sequence
45     25      5872      9      AR101810 Sequence

```

ALIGNMENTS

```

RESULT 1
LOCUS AR029499 1041 bp DNA
DEFINITION Sequence 7 from patent US 5859335.
ACCESSION AR029499
VERSION AR029499.1 GI:5941472
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1041)
AUTHORS Patton,D.Andrew,
TITLE Enhanced biotin biosynthesis in plant tissue
JOURNAL Patent: US 5859335-A 7 12-JAN-1999;
FEATURES
Source Location/Qualifiers
1..1041
/organism="unknown"
BASE COUNT 262 a 273 c 305 g 201 t
ORIGIN

```

```

Query Match 100.0%; Score 25; DB 9; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.00011;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 ggatctgctgttgaagcagcag 25
|||||
Db 63 GGATCTGCTGTTGAAGCCAGCAG 87

```

RESULT 2

```

AR034916 AR034916 1041 bp DNA PAT 29-SEP-1999
LOCUS AR034916
DEFINITION Sequence 7 from patent US 5869719.
ACCESSION AR034916
VERSION AR034916.1 GI:5950521
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1041)
AUTHORS Patton,D.A.
TITLE Transgenic plants having increased biotin content
JOURNAL Patent: US 5869719-A 7 09-FEB-1999;
FEATURES
Source Location/Qualifiers
1..1041
/organism="unknown"
BASE COUNT 262 a 273 c 305 g 201 t
ORIGIN

```

```

Query Match 100.0%; Score 25; DB 9; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.00011;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 ggatctgctgttgaagcagcag 25
|||||
Db 63 GGATCTGCTGTTGAAGCCAGCAG 87

```

```

RESULT 3
LOCUS A11530 1084 bp DNA PAT 09-FEB-1994
DEFINITION Bior gene of E.coli with primers.
ACCESSION A11530
VERSION A11530.1 GI:490218
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Escherichia coli.
REFERENCE 1 (bases 1 to 1084)
AUTHORS Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
JOURNAL Escherichia.
FEATURES
Source Location/Qualifiers
1..1084
/organism="Escherichia coli"
/db_xref="taxon:562"
24..1064
/gene="BIOB"
24..1064
/gene="BIOB"
24..1064
/gene="BIOB"
/codon_start=1
/transl_table=1
/protein_id="CAA00965.1"
/db_xref="GI:490219"
/translation="MAHRRPWTLSQVTELFKPLDLLEFAQVYHROHEDDPROVYST
LSTKTCACPEDCKYCPQSSRYRTGLAEERLMEVEVLESARAKAGSTRFCMGAW
KNPERMDPYLEQWQVQKAMGLACWTGLTSESOARLANAGLDYNNHNDTSPF
YGNITTRTYOERLDLEKVRDAGIKVCSGIVGLGETVDRAGLLLOLANLPTPES
VPINMLVKVGTPLANDVDADFDFTITIVARIMPTSVRLSAGREOMNEQTOAMC
FMAGANSIFYGCKLITTPNEDEKDLQLPFKGLNQQRAVLAGNDEQQRLEQALMT
PDDDEVNNAAL"

```

```

BASE COUNT 271 a 286 c 318 g 209 t
ORIGIN

```

```

Query Match 100.0%; Score 25; DB 9; Length 1084;
Best Local Similarity 100.0%; Pred. No. 0.00011;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ggatctgctgttgaagcagcag 25
|||||

```


ACCESSION J04423
 VERSION 7.8-diamino-pelargonic acid aminotransferase;
 KEYWORDS 7-keto-8-amino-pelargonic acid synthetase; bioA gene; bioB gene;
 SOURCE bioC gene; bioD gene; bioF gene; biotin synthetase; dethiobiotin
 ORGANISM synthetase.
 Escherichia coli (strain K-12) DNA.
 Escherichia coli
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Escherichia.
 1 (bases 1 to 5793)
 REFERENCE Otsuka,A.J., Buoncristiani,M.R., Howard,P.K., Flamm,J. and
 AUTHORS Johnson,O.
 TITLE The Escherichia coli biotin biosynthetic enzyme sequences
 JOURNAL J. Biol. Chem. 263, 19577-19585 (1988)
 MEDLINE 89066784
 COMMENT Draft entry and computer-readable sequence [1] kindly submitted by
 A.Otsuka, 09-NOV-1988.
 FEATURES
 SOURCE Location/Qualifiers
 1..5793
 /organism="Escherichia coli"
 /strain="K-12"
 /db_xref="taxon:562"
 complement(98..574)
 /note="ORF 1"
 /codon_start=1
 /transl_table=11
 /protein_id="AA023513.1"
 /db_xref="GI:455168"
 /translation="MKLISNDLRDGLPHRHVFNCGCYDGNISPLAMDVPACTK
 SFVYCYDPDAPRTGSGWMMVYVNLPAIDTVYLQGGSGVYAMPDVLQTRDFGTG
 YDGAAPKRGTHRTFTVYVHLDIDVDGASAMVGFVNHSLASATIMS"
 complement(633..1925)
 /gene="bioA"
 complement(633..1925)
 /gene="bioA"
 /codon_start=1
 /transl_table=11
 /product="7,8-diamino-pelargonic acid aminotransferase"
 /protein_id="AA023514.1"
 /db_xref="GI:457106"
 /translation="MTTDLAFDORHIMHPYTSPLPYVPVSAEGCLILDSGR
 LVDMSSWMAAIGHYVNPOLNAAMKSOIDMSHVMGGITTHAPATLCKRYAMSGRN
 ALBCVFLADSGVAEVAEMKALQYMAKAGORPLTRNGHGTTCAMSCDDN
 SMHSIMKCYLPENLAPAPQSRMDGEMDEDMGRALMAHHEHETAAVILETVGA
 GGMARHPEWIKRIKTRICDREGILLIADETATGFRGKLFACEHEIAVILICLGA
 LTGGWTLNLTATLTREVAETISNGEAGCEMHGPTFNGNPLACAAANASLAIESGDMO
 QQVADIEVQLREQLAPARDAEMVADVIVLGAIGVETTHPVNMAALQKFEVFGVWIR
 PFGKLIYLMPPYIILPQOLRLTAAVNRAVQDETFFCQ"
 2012..3052
 /gene="bioB"
 2012..3052
 /gene="bioB"
 /codon_start=1
 /transl_table=11
 /product="biotin synthetase"
 /protein_id="AA023515.1"
 /db_xref="GI:145425"
 /translation="MAHRRRWLTISOYTELEFKPLDLLEAOQVHRQHPDPROVST
 LLSIKTGACEDCKYCPQTSRKYTGLEARLEMEVEVLSARAKANAAGSTRCMGAAM
 KNHERMDPYLEDMOVGVKAMGLACMTGLTSESQAQALNAGLDYYNHNIDTSPEP
 YGNIITTRTYOERLDLEKRVDAIGIVCGSGIVGEGYKRBAGLLLOLAN.PTPSES
 VPINMLVYKGTPLADNDVDADFRTITAVARIMMPTSYVRLSAGREDMNOCOTAMC
 PMAGANSYFGCKILTTNPEEDKQLQLFRKGLNPGQTAVALAGNEEOQRLQALMT
 PPTDEYNAAL"
 3049..4203
 /gene="bioF"
 3049..4203
 /gene="bioF"
 /codon_start=1
 /transl_table=11

/product="7-keto-8-amino pelargonic acid synthetase"
 /protein_id="AA023516.1"
 /db_xref="GI:145426"
 /translation="MSWDERKINAALDARRAADALRRRYPAVAGAGRWLVADROLYNF
 /transl_table=11
 /protein_id="AA023517.1"
 /db_xref="GI:145427"
 /translation="MATVKKQAIIAAFGRAAAHYBOHADLQROSDALLAMLPORRY
 HYLDAGCGGMSRHRERHRAQVTLADLSPMLVOAROKDADHYLAGDITSLPLATA
 TFDLWASNLAVOMCGNLSTALRELYRVVRKGVVAFVTLVGGSLPERHOAOVDERP
 HANRFLPDEIRDSLNGVYHOHIDPITLMPEDALSAMRSKIGATLHGRDRRL
 TRSOLQRLQIAMPQOQGRPLTYHLFLGIYARE"
 4938..5597
 /gene="bioD"
 4938..5597
 /gene="bioD"
 /codon_start=1
 /transl_table=11
 /product="dethiobiotin synthetase"
 /protein_id="AA023518.1"
 /db_xref="GI:145428"
 /translation="MSKRYFYVGTDEVEKTVASCALLQAAKRAGYRTAGKPYVAGS
 EKTPEGLNSDVALIQRNSLQIDYATVATPFAEPSPHIIISOEGPISLVMSAG
 LRALRQADWVAVLEGAGCMFTPLSDTFFFAWMVQEOLEPVLVYVGGKGINHMLTA
 QVTHAGLTLGAVANDVTPGKRRAEEMTLTTRIPRCWERSBPGLQIKMKOPES
 T"
 BASE COUNT 1363 a 1554 c 1631 g 1245 t
 ORIGIN 4626 bp upstream of HpaI site; 18 min on K-12 map.
 Query Match 100.0%; Score 25; DB 2; Length 5793;
 Best local Similarity 100.0%; Pred. No. 0.00011;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ggaatcgtcgtttgaagcgacagc 25
 ||||||||||||||||||||
 DB 2074 ggatctgctgtttgaagcgacagc 2098
 ||||||||||||||||||||
 RESULT 6
 A38246 5872 bp DNA PAT 05-MAR-1997
 LOCUS Sequence 1 from Patent WO9408023.
 DEFINITION A38246
 ACCESSION A38246
 VERSION A38246.1 GI:2294844
 KEYWORDS
 ORGANISM
 Escherichia coli.
 Escherichia coli
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Escherichia.
 1 (bases 1 to 5872)
 REFERENCE Birch,O., Brass,J., Fuhrmann,M. and Shaw,N
 AUTHORS BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
 TITLE Patent: WO 9408023-A 1 14-Apr-1994;
 JOURNAL LONZA AG (CH)
 COMMENT
 other publication PL 308301 950724
 other publication CA 2145400 940414
 other publication AU 4820293 940426
 other publication HU 71781 960228
 other publication SK 42095 951108
 other publication CZ 950809 950913
 other publication FI 951547 950331

Other publication JP 8501694F 960227.
 Location/Qualifiers
 1. 5872
 /organism="Escherichia coli"
 /strain="DSM498"
 /db_xref="taxon:562"
 /clone="PBO30A-15/9"
 1. 96
 /function="PROMOTER PTAC"
 /evidence=experimental
 23. 28
 45. 49
 /standard_name="PROMOTER PTAC"
 /evidence=experimental
 105. 119
 /standard_name="BIOB RBS NO.9"
 /evidence=experimental
 117. 1157
 /gene="BIOB"
 117. 1157
 /gene="BIOB"
 /codon_start=1
 /transl_table=11
 /number=1
 /evidence=experimental
 /product="BIOTIN SYNTHASE"
 /protein_id="CAA02325.1"
 /db_xref="GI:2294845"
 /translation="MAHPRWLTLSQVTELFEPDLDDLEFAQGVHROHDPPOVYST
 LSTKTCGPEDECKYCPSSRYKTLGAEFLMEBOVLESAKAKAAGSTFCMGAM
 KNHERMPYLEOMOGYKMGLEACMTLCTTSSAOARLANGLDYNNINDTSPF
 YGNTITRTYQERLDITLKYRDKGKICVSGIYGLSETYKDRAGLLLOLANPTPEPS
 VPINMLVKVGTPLADNDVDADF1RTTAVARIMPTSTVRLSAGREBONNETOAMC
 FMAGANSIFYGCKLITTPNDEBDKDLDFKRLGINPOQAVIAGDNEQOQRLQALMT
 PTDVEYNAAL"
 2284. 2297
 /standard_name="BIOC RBS"
 2295. 3050
 /gene="BIOC"
 2295. 3050
 /gene="BIOC"
 /function="INVOLVED IN PIMETOYL-COA SYNTHESIS"
 /codon_start=1
 /transl_table=11
 /number=3
 /product="PROTEIN"
 /protein_id="CAA02326.1"
 /db_xref="GI:2294846"
 /translation="MATVNRQATIAAFGRAAHYEHOADLOROSADALLAMLPOKRYT
 HYLDAGCGPGMSHMRPERIAQVYALDSPKMLVOAROKDAADHYLAGDIESPLATA
 TFDLWNSLAWQMGNGSTALRELRYVRKGVAFITLVQGSLEPELQAMQAVDERP
 HANFELPPDEIEQSLNGVYOHHTOPTLWDDDLMSRLSKIGATVHLHEGRPRL
 TRSOLRLOLAMPQOGRPLTVHFLGVIARE"
 3742. 3752
 /standard_name="BIOA RBS"
 3750. 5039
 /gene="BIOA"
 3750. 5039
 /gene="BIOA"
 /standard_name="S-ADENOSYL-L-METHIONINE:
 8-AMINO-7-OXONONANOATE AMINOTRANSF."
 /EC_number="2.6.1.62"
 /codon_start=1
 /transl_table=11
 /number=5
 /evidence=experimental
 /product="DAPA SYNTHASE"
 /protein_id="CAA02327.1"
 /db_xref="GI:2294847"
 /translation="MTTDDLAFDORHIMHTPTSKTSPVPYVVSABCECLLSDGR
 LVDGSSWMAA1HGNHPOLNAAMKSQIDAMSHWFGITHPA1ELCRKLVAAVTPOP
 LECVFLADSGSYAVAMKMALOYWQAGEARORFLTFRNQYHGDITFGAMSVCDPDS
 MHSIMKGYLPENTLFAPOSIMDGEWDERDMVGFARLMAAHRHEITAVIIEPIVQAGS

GMRYHPPEWLKRIKICDREGILLIADIATGRTGKLFACENAEIAPDILIGRAL
 TCGMTLSATLITREVAETISNGEAGCFMGPTEMGNPPLACAAANASLAIIESGDMQO
 QVADIEVOLREQLAPARDAENVADYRVLGAYETHPVPMMAALOKFVEQGVWIRP
 FGKLIYMPPIILLPQOLRILTAAVNRAVODETFEQ"
 5088. 5100
 /standard_name="ORF1 RBS"
 5098. 5574
 /gene="ORF1"
 5098. 5574
 /gene="ORF1"
 /function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
 /codon_start=1
 /transl_table=11
 /number=6
 /evidence=experimental
 /product="PROTEIN"
 /protein_id="CAA02328.1"
 /db_xref="GI:2294848"
 /translation="MKLISNDLRDGLKLPHRHVFNGMGYDGDNISPHLAMDVPAKTK
 SPVVCYDPAVGTSGMHWVYVNLPAITRVLPQGFSGVLAMPDGYLQTRTDFGKTG
 YDGAAPRGERTHRYFTYHALDIERIDVDEGASGAMGFVHPHSIASITAMFS"
 5583. 5605
 5583. 5644
 /standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
 TERMINATOR"
 1552 c 1695 g 1307 t

BASE COUNT 1318 a 1552 c 1695 g 1307 t
 ORIGIN
 Query Match 100.0%; Score 25; DB 9; Length 5872;
 Best Local Similarity 100.0%; Pred. No. 0.00011;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcctgccttttgaagcgagcag 25
 |||||||
 Db 179 GGATCTGCTGTTGAAGCGCAGCAG 203

RESULT 7
 A38251 5872 bp DNA PAT 05-MAR-1997
 LOCUS
 DEFINITION Sequence 6 from Patent WO9408023.
 ACCESSION A38251
 VERSION A38251.1 GI:2294849
 KEYWORDS
 SOURCE
 ORGANISM
 Escherichia coli.
 Escherichia coli.
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Escherichia.
 1 (bases 1 to 5872)
 Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.
 BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
 Patent: WO 9408023-A 6 14-APR-1994;
 IONZA AG (CH)
 Other publication PL 308301 950724
 Other publication CA 2145400 940414
 Other publication AU 4820293 940426
 Other publication HU 71781 960228
 Other publication SK 42095 951108
 Other publication CZ 9500809 950913
 Other publication FI 951547 950331
 Other publication JP 8501694F 960227.
 location/Qualifiers
 1. 5872
 /organism="Escherichia coli"
 /strain="DSM498"
 /db_xref="taxon:562"
 /clone="PBO30A15-9"
 1141. 1156
 /standard_name="BIOB RBS"
 1154. 2308
 /gene="BIOB"
 1154. 2308

FEATURES
 source
 RBS
 gene
 CDS

```

/gene="BioF"
/standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
/EC_number="2.3.1.47"
/codon_start=1
/transl_table=11
/number=2
/evidence=experimental
/product="KAPA SYNTHASE"
/protein_id="CA02329.1"
/db_xref="GI:2294850"
/translation="MSWQEKINALDARRADALRRHYPAVAGRWLVADPROYIN
SSNDYLGSHHPOTIRAMOGAEQFGSGSGSGSYVHOALEEELAEWLCYSA
LLFISGFANQAVIAAMAKEDRIIARLSHSLLEASLSPSQLREAFANDVHLAR
LLASPCGQOMVTEGVFSDGDSAPLAEIOVQHNGLMWDAGHGTGVEQNG
SCHLQVAPPELLVTEGKFGVSGAAVCSIVADYDLOFARHLIVSTMPAQOAL
RASLAVIRSEGDARREKLAALITFRAGVODLPFTLADSCSAIQPLIVGDSRALQL
AEKLRQGCWYAIRPTVPAGTARLRLTLAAHEMODIRLLEVLHGNC"
3030..3045
/standard_name="BIOB RBS"
3043..3753
/gene="BIOB"
3043..3753
/gene="BIOB"
/standard_name="DETHIOBIOTIN SYNTHASE"
/EC_number="6.3.3.3"
/codon_start=1
/transl_table=11
/number=4
/evidence=experimental
/product="DTB SYNTHASE"
/protein_id="CA02330.1"
/db_xref="GI:2294851"
/translation="MSKRYVNGTDREVGKTVASCALLQAKAGRTGKYPVAGS
EKREPCARNSDALALORNSLOLDYATVNPYFAEPTSPHIIISQEGRPESLWAG
LRLEQOQADWLVLEGAGWFTPLDFTFADWVTOQLPVILVGVKLCINHAMLTA
QVTOHAGLTLTAGVANDVTPPKRAHAEYMTTLRMIPAPLGEIPWLAENPENAAATGK
YINLAEVASTLIGFTSRL"
BASE COUNT      1318 a      1552 c      1695 g      1307 t
ORIGIN
Query Match      100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.00011;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1      9gacctgctgttgaagcgacag 25
      |||||||||||||||||||
Db      179 GGATCTGCTGTTTGAAGCGACGAG 203
RESULT      8
LOCUS      A93674      5872 bp      DNA
DEFINITION      Sequence 1 from Patent EP0798384.
ACCESSION      A93674
VERSION      A93674.1 GI:6741862
KEYWORDS
SOURCE      Escherichia coli.
ORGANISM      Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE      1 (bases 1 to 5872)
AUTHORS      Birch, O. and Brass, J.
TITLE      Biotechnological method of producing biotin
JOURNAL      Patent: EP 0798384-A 1 01-OCT-1997;
LOCUS AG (CH)
FEATURES
source
1..5872
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/clone="PB030A-15/9"
1..96
promoter

```

```

-35_signal
-10_signal
RBS
gene
117..1157
/gene="BIOB"
117..1157
/gene="BIOB"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="BIOTIN SYNTHASE"
/protein_id="CAB69590.1"
/db_xref="GI:6741863"
/translation="MAHRPWTLSQVTELEKPLDLLFEAQOYVROHFDPROYVST
LISITGACPECKYCPQSSRYKGTLEARLMEVEQVLESARAKAGSTPCGAAW
KNRHRDMPYLEOMVOGYKAMGLGTLTSBOQARLNAGIDYTNHLDPSER
VGNITITTRYQERLDTLKVADAGIKVCSGIGVIGETVKDQAGLLQLANLPTPES
VFIMNLVKGFTPLADNDVDAFDFIRLIAARIMPTSVRLSAGRQONQOAMC
FMAGANSIFYGCKLLTPNPEDKDLQLEFRLLGINPQOTAVIAGDNEQQRLEQALMT
PDIDEYNAAL"
2284..2297
/standard_name="BIOC RBS"
2295..3050
/gene="BIOC"
2295..3050
/gene="BIOC"
/function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
/transl_table=11
/codon_start=1
/product="PROTEIN"
/protein_id="CAB69591.1"
/db_xref="GI:6741864"
/translation="MATVYNKOALAAFGRAAHYEQHADLQROSDALLMLPQRYT
HYLDAGCGPQWMSHMRERHAQVATLALDLPMLVQAOAKDAADHYLAGDIESPLATA
TFEDLAPNDELAQWQCNLSTALRELYRVVRPGVAAFTTIVGSLPELHOAMQVDESR
HANRLPLMDEIEOSLNGVYOHYIOTITIMPDDALSMBSLKGATGHHLEGNRRLT
TRSOLOROLAMPOQGRYPLTHLFLGVIARE"
3742..3752
/standard_name="BIOA RBS"
3750..5039
/gene="BIOA"
3750..5039
/gene="BIOA"
/standard_name="S-ADENOSYL-L-METHIONINE-8-AMINO-7-
OXONONANOATE AMINOTRANSF."
/EC_number="2.6.1.62"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="DAPA SYNTHASE"
/protein_id="CAB69592.1"
/db_xref="GI:6741865"
/translation="MTTDLAFDORHIMPTSTPLVYPVVASBGCCELLISDGR
LVDSQSSWMAIHGYNHPOLNAAMKSOIDMSHVMGGITTHAAIECRVLVMTPOP
LECVFLADSGSAVEYVNAALQYQWAKGARGRFLFRNGYGDFFGASVYCDPNS
MHSIMKGYLPENLFAPAPQSRMDGEMDERMVGAFARLMAHREHIAVIEPIVQAG
GMRMYHPEWLKRIKICDREGILLADEITGFGRGKLPACFHAELADPILGKAL
TGGMITLSATLTLREVAETISNGEAGCPFMGPNPLACAAANASIALTESGDMQO
OAGDIEVOUTREOLAPARDAEMVADYRLGATIGVETTHPPNMAALOKFVEQGVWIRP
FGKILYIPPYIILPQOLORTAIVNAVODETFEQ"
5086..5100
/standard_name="ORF1 RBS"
5098..5574
/gene="ORF1"
5098..5574
/gene="ORF1"
/function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"

```

[illegible]

source 1..5872
 /organism="unknown"
 BASE COUNT 1318 a 1552 c 1695 g 1307 t
 ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 5872;
 Best Local Similarity 100.0%; Pred. No. 0.00011;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggaatgcgttggaagcagcag 25
 |||||||||||||||||||||
 Db 179 GGATCGCTGTTGAAGCGCAGCAG 203

RESULT 12
 AEO00180 11022 bp DNA ECT 01-DEC-2000
 LOCUS Escherichia coli K12 MG1655 section 70 of 400 of the complete
 DEFINITION genome.
 ACCESSION AEO00180 U00096
 VERSION AE000180.1 GI:1786988
 KEYWORDS
 SOURCE
 ORGANISM Escherichia coli K12.
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Escherichia.

REFERENCE 1 (bases 1 to 11022)
 AUTHORS Blattner, F.R., Plunkett, G., III, Bloch, C.A., Perna, N.T., Burland, V.,
 Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F.,
 Gregor, J.J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J.,
 Mau, B. and Shao, Y.
 The complete genome sequence of Escherichia coli K-12
 Science 277 (5331), 1453-1474 (1997)

TITLE JOURNAL
 MEDLINE 9742617
 PUBMED 9278503
 REFERENCE 2 (bases 1 to 11022)
 AUTHORS Blattner, F.R.
 TITLE Direct Submission
 JOURNAL Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
 University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
 Email: ecollegenetics.wisc.edu Phone: 608-262-2534 Fax:
 608-263-7459
 3 (bases 1 to 11022)
 REFERENCE 3
 AUTHORS Blattner, F.R.
 TITLE Direct Submission
 JOURNAL Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
 University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
 Email: ecollegenetics.wisc.edu Phone: 608-262-2534 Fax:
 608-263-7459
 4 (bases 1 to 11022)
 REFERENCE 4
 AUTHORS Plunkett, G. III.
 TITLE Direct Submission
 JOURNAL Submitted (13-OCT-1998) Laboratory of Genetics, University of
 Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
 This sequence was determined by the E. coli Genome Project at the
 University of Wisconsin-Madison (Frederick R. Blattner, director).
 Supported by NIH grants HG00301 and HG01428 (from the Human Genome
 Project and NCHGR). The entire sequence was independently
 determined from E. coli K12 strain MG1655. Predicted open reading
 frames were determined using Genemark software, kindly supplied by
 Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
 30333 (e-mail: mark@ember.gatech.edu). Open reading frames that
 have been correlated with genetic loci are being annotated with CG
 site Nos., unique ID nos. for the genes in the E. coli Genetic
 Stock Center (CGSC) database at Yale University, kindly supplied by
 Mary Berlyn. A public version of the database is accessible
 (http://cgsc.biology.yale.edu). Annotation of the genome is an
 ongoing task whose goal is to make the genome sequence more useful
 by correlating it with other data. Comments to the authors are
 appreciated. Updated information will be available at the E. coli
 Genome Project's World Wide Web site
 (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and

its annotations are periodically updated; this is version M54. No
 sequence changes. Annotation updates: updated gene identifications
 and products; all new functional assignments courtesy of Monica
 Riley; added promoters, protein binding sites, and repeated
 sequences described in reference 1. The unique numeric identifiers
 beginning with a lowercase 'b' assigned to each gene (protein- or
 RNA-encoding) are now designated as gene synonyms instead of
 labels. This should allow them to be searched for in Entrez as gene
 names.

FEATURES

source
 Location/Qualifiers
 1..11022
 /organism="Escherichia coli K12"
 /strain="K12"
 /sub_strain="MG1655"
 /db_xref="taxon:83333"
 <1..130
 /note="REP (repetitive extragenic palindromic) element;
 contains 4 REP sequences"
 complement(147..1430)
 /gene="ybhc"
 /note="b0772"
 complement(147..1430)
 /gene="ybhc"
 /function="orf; Not classified"
 /note="f427; 98 pct identical to fragment YBHC_ECOLI
 SW:P46130 (300 aa) but contains 127 additional C-terminal
 residues"
 /codon_start=1
 /transl_table=11
 /product="putative pectinesterase"
 /protein_id="AAC73859.1"
 /db_xref="GI:1786989"
 /translation="MNTFSVSRLLALAFVTLTACSGSTPPDRSPDPTAGTSRRP
 LSAFAONFNDQHYVPSLTPGAAMNPSPITLPQPPVNPACTGCTHTTTOAAD
 AAIKRTKROYIAVMPGECYTVPAAPGCTITLTGTEKPTDKTGLSDGMSFA
 DMRHVNPNGKIMPKPAWMTDSCQKRSQICVLSAVFWSQNNGLQNLQNTENT
 LQSDVDAGNHPVALRTDQDOVINNNVITIGKQNTFFVTNSGVQNRLETNRQPTLV
 NSYLEGDVIVSGAGVAVDNTEFRVNVNSRTQDEAVYVAPATLSNIYGFIAVSRN
 AFGVDQVQLGRSLVDANTNGQVVIKDSAINEGFNTAKPMADAVISNRPFAGNTGSDV
 DNDEIQRNLNDNTNRMVEYNNGVSGKVAEAK"
 complement(1507..1535)
 /note="factor Sigma70: predicted +1 start at 806574"
 complement(1582..2058)
 /gene="ydhb"
 /note="b0773"
 complement(1582..2058)
 /gene="ydhb"
 /function="orf; Unknown"
 /note="f158; 99 pct identical to YBHC_ECOLI SW: P12994"
 /codon_start=1
 /transl_table=11
 /product="orf, hypothetical protein"
 /protein_id="AAC73860.1"
 /db_xref="GI:1786990"
 /translation="MKLISNDLRGDKLPHRHVPNGXGCDNISPLADVDPAK
 SFVVTGYPDPAQSGGHHWYVLPADTPVLQGRSGGIYAMPDGLQTRTPGTG
 YDGAAPKRGKRNRTFTVHALDRIIDVDGASGAMGFVNHSLASASITAFMS"
 complement(2108..2124)
 /note="central position to predicted promoter:85..5"
 /bound_moiety="Rna5 predicted site"
 complement(2117..3406)
 /gene="b10A"
 /note="b0774"
 complement(2117..3406)
 /gene="b10A"
 /EC_number="2.6.1.62"
 /function="enzyme; Biosynthesis of cofactors, carriers;
 Biotin"
 /note="f429; 100 pct identical to B10A_ECOLI SW: P12995"
 /codon_start=1
 /transl_table=11
 /product="7,8-diaminopelargonic acid synthetase"
 /protein_id="AAC73861.1"

Query Match

100.0%; Score 25; DB 1; Length 11022;

/db_xref="GI:1786991"
 /translation="MTTDLAFDQRIHWPYTSMTPLPYVPSAEGCELLISDGR
 LVDMSSWMAAIGHYNHNPOLNAAMKSOIDAMSHVMGGLTHAFALCKRLVAMPQP
 LECVPLADSGSAVEAVKMAIOYMOAKGARORFLFRNGYGDFTGMSVCDPNS
 MSLIMKGYLPENFAPAPOSRMDGEMDERMVGAFRLMAHREIAVILIEPVGAG
 GMMKRYPEWIKRTIKCDRCGLILIDELATGCRGACFACCHADILICLGLAL
 TGGTMTLSATLTTRVAETTSNGEAGFMHGPFGNPLACAAANSIALILEGDMQ
 OVADIEVLEQLOAPADAEVAVDVLGIVETTHPVNMAALOKFVEGCVWLRP
 FGKILYIMPPYILLPOLQRLTAVNNAVDDEFPCO"
 complement(2193..2221)
 /gene="bioA"
 /note="factor Sigma70; predicted +1 start at 807260"
 3411..3450
 /note="central position to bioB promoter: -20"
 /bound_moiety="BioB documented site"
 3411..3450
 /note="central position to predicted promoter: -20; genetic
 evidence for the site"
 /bound_moiety="BioB predicted site"
 complement(3411..3450)
 /note="central position to predicted promoter:50"
 /bound_moiety="BioB predicted site"
 complement(3411..3450)
 /note="central position to bioA promoter:50"
 /bound_moiety="BioB documented site"
 3413..3441
 /note="factor Sigma70; promoter bioB; documented +1 at
 808525"
 complement(3447..3473)
 /note="factor Sigma70; promoter bioA; documented +1 at
 808515"
 3493..4533
 /gene="bioB"
 /note="b0775"
 3493..4533
 /gene="bioB"
 /EC_number="2.8.1.-"
 /function="enzyme; Biosynthesis of cofactors, carriers:
 BioIin"
 /note="0346: 99 pct identical to BIOB_ECOLI SW: P12996"
 /codon_start=1
 /transl_table=1
 /product="bioIin synthesis, sulfur insertion?"
 /protein_id="AAC73862.1"
 /db_xref="GI:1786992"
 /translation="MAHRPMTLSQVTELFEKPLDLLFEAQVYHROHFDEROVAVST
 LLSIKTACPEDCKYCPSSRYKTGLFAERLMEYOVLSEAKAKAGSTFPCGAMV
 KNPHRMDPYLEOMVGVKAMGLACQMTLGLSQAORLANAGIDYNNHLDSPES
 YGNITTTTYOERDLTEKRVADGKIGTISGQVXRAGILLOLANLPTPPES
 VPINMLVVKGTPLADNDVDADFIRITIAVARIMPTSYRLSNGRQONQOAMC
 FMAGANSIFYGCKLITTPNPEKDLQLEKRLGIMPOTAVLAGDNEQORLEQALMT
 PDUDEYNAAL"
 4530..5684
 /gene="bioF"
 /note="b0776"
 4530..5684
 /gene="bioF"
 /EC_number="2.3.1.47"
 /function="enzyme; Biosynthesis of cofactors, carriers:
 BioIin"
 /note="0384: 100 pct identical to BIOF_ECOLI SW: P12998"
 /codon_start=1
 /transl_table=1
 /product="8-amino-7-oxononanoate synthase"
 /protein_id="AAC73863.1"
 /db_xref="GI:1786993"
 /translation="MSMOEKINALDARRAADALRRYPVAGAGRWLVADROYLNF
 SNDYLSHHPOLIRANQOGEORCSGSYGVVAGQALEEELAEGLSRA
 LFIISGFANQAVIAAMAKEDRIADRLSHSLLEASLSLSQIRRAHNDVTHLAR
 LLASPCGOQVNTGEVFSMDGSAFLAEIQVTOOHNGMGLVDDHSGTYGEGRG
 SCMLQKVPKELLVTFEGKFGVGAALVCSSTVADYLQFAHRLIYTSMPAQAOL

Best Local Similarity 100.0%; Pred. No. 0.00012;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggaatctgctgtttgaagcgagcag 25
 |||
 Db 3555 GGATCTGCTGTTTGAAGCGACGACGAG 3579

RESULT 13

AE004192

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

```

/gene="VC1112"
/Note="similar to GB:J04423 SP:P12996 PID:145425
PID:490219 GB:U00096; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=1
/product="biotin synthase"
/protein_id="AA094271.1"
/db_xref="GI:9655583"
/translation="MEVHNMTVAEVAKALDLKPPMDLLEAQQVYHRLHPHNVOVST
LISITGACPEDCKYCPQSAHYRIDVDERLEMEYERVIDAOKAKNGSGTRFCGAM
KNPKRDMPLKEMIRGKMDGLTCTGTLGMLTPDOAQLAQAQIDYNNHLDTSPE
YGNITTTTODRLDPLSHVDACMKICSGITOMGISTNDAGLLEVLNLPHPSS
VPIWMLYKVTPELEQVDDVDFPDLVLAARIMPKSAVRSAGREKKNEDQALC
PMAGANSTFEGCKLITTPNPAEDSDMLLFKRLGINROVAKPDEITENELDRVER
VAARP7ASDLEFYDAAL"
2533..3687
/gene="VC1113"
2533..3687
/Note="similar to GB:J04423 SP:P12998 PID:145426
PID:490225 GB:U00096; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=1
/product="8-aminio-7-oxononanoate synthase"
/protein_id="AA094272.1"
/db_xref="GI:9655584"
/translation="MTLRFKSRIEQLALNREHSHLTPRLLDLRSAGGELIQDRTYL
NESSNDYLGANDALVQAMOTGLARFGAGSAPVATGSSAHAELEHLCMIGLE
RAVLFESSGFSNOMALFTLEKEDLLQDKLNHSLMEAGMLSPATMKRKNDTHL
ROLHEGOSNLVTEGVFSDMGDCAPLAOLATLAKORAWLMDVDAHGICVLDGDCGS
COAGIHPEITIVTFEKGAFGAGAAVLCDAHVDGYLTOPARHHYVSTAMPAPAAHAT
HAARMIOGSMREROLAEIACDEQCRNTPGVATOTPIKPMILSSDSALASHL
KOCGIWTSATRPPTVPGSARLRTITTAHTQATRLSEOLKHMEDT"
3549..4487
/gene="VC1114"
3549..4487
/Note="similar to GB:J04423 SP:P12999 PID:145427
PID:490221 GB:U00096; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=1
/product="biotin synthesis protein Bioc"
/protein_id="AA094273.1"
/db_xref="GI:9655585"
/translation="MGECHSTTHRACRQCAFTHDNGAYPSADPPLIGAVKTCDEG
HMSMTATLCEVELKDKSALIAAFGAARHDOHAOFORVGRLLQKMPSCGLGLVL
DLGGTGTFSALLREKGAQVVCADISHAMELOAKORCGDEGMSYQIADAEOLPFAAS
EDMVFSLSLAQWCEDELPLSEIRRVKPHGQAFSTLSDSLFELQAMRSVDHHRH
INOFISINDVKIALQAGCSQHHLDLAITVWEAFMLRDLKIGAHVSGRSRGL
ISRRLAKVELAAVQSPRNOGLVPAITYQVCLGYIHR"
4463..5179
/gene="VC1115"
4463..5179
/Note="similar to GB:S68059 SP:P13000 PID:490223
PID:145428 PID:460901; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=1
/product="dehydrobiotin synthetase"
/protein_id="AA094274.1"
/db_xref="GI:9655586"
/translation="MFRGYSMPRANAIFLAGTDDVKGTYASKAIIQALATNIAITG
KVAAGSDKTEGVRNSDALHMKAAATVMDPEYDPAIVPTSPHAAHKNITIG
VALLSNKLSKHKONAEIVYEGAGCRVPPTSDDLSTWVQERLPVILVIGIKGCL
SHAITFAEIRADGELVGMVIANRINPGEHVAELIEHLERGLTGPRLGEIPEYMKAK
ROELKFIQIDHLLLEPDVYA"
complement(5261..5470)
/gene="VC1116"
complement(5261..5470)

/gene="VC1116"
/Note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AA094275.1"
/db_xref="GI:9655587"
/translation="MSRRNLITNKCKHMAEASHYQMAIDLCCHLGISEQEARQGL
EKEQRTLOOQIYDTQSHMGLSAEK"
5697..6560
/gene="VC1117"
5697..6560
/Note="similar to GB:M58470 SP:P23894 PID:146412 GB:U00096
PID:1736470; identified by sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="heat shock protein HspX"
/protein_id="AA094276.1"
/db_xref="GI:9655588"
/translation="MKRILLPLATNLAVVIVSVYLVNTVAVTGMQPSLSGLVMAA
VFGFGARFISILMSKSMALRSQGVYIDTPPREMHWLLETVRQANQAGIGPTVAI
YAAPDMNAFTGAKRDSILVAYSTGLHNMTRDEEAVLAHEVSHIANGDVYTMIMQ
GVNFTVFIFLSRFIANIVASRUSEGEGSNMAYVGVSMVLELVFGFLASTITMYSR
HREPFADAGAAQLVGRHKMIALERLKMGOESHLGSSMARGITGRSLSLMTMHP
LEKRIALARNM"
6770..7198
/gene="VC1118"
6770..7198
/Note="similar to PID:1653608; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=1
/product="transcriptional regulator, putative"
/protein_id="AA094277.1"
/db_xref="GI:9655589"
/translation="MSIMDVQVHVAQFQQLDKSOLHRIETYPHDVFEEDAAHREGEF
DALYVFLNLYONVMTCTFTIHEQVAVNEGAFVWTMHLRHPKLAKGBOVDYKGVSHL
HFAEKVTHHRYFDMGEMLEQLEQVIRAIKRRIGQ"
7195..7920
/gene="VC1119"
7195..7920
/Note="similar to SP:P25970; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=1
/product="oxidoreductase, short-chain
dehydrogenase/reductase family"
/protein_id="AA094278.1"
/db_xref="GI:9655590"
/translation="MTAVFTIGATSGIGKQLAIDYAKQGVYIACGNQSVLDSLHTQ
YANIFPLAFVTDHPTKALALQPCPELMTIINAGCEYIDGKMDVIMAVFNN
VLGAAVAVTEGLOPHLSCGHRVAIYASIELALPRAYGASAAVYIARTIOLDMR
PLGEVNTIPEGFVATPLTDNTPFAMNMITVERAAQETKAGIARVSQLPFRRTFW
LIRLLGALPYAMQGRVLRRLIKA"
7931..9211
/gene="VC1120"
7931..9211
/Note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=1

Query Match 100.0%; Score 25; DB 1; Length 12891;
Best Local Similarity 100.0%; Pred. No. 0.00012;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ggatctgctgttgaagcgagcgagc 25
DB 1559 ggactctgctgttgaagcgagcgagc 1583

```

```

RESULT 14
LOCUS AE005258 13501 bp DNA BCT 21-MAR-2001
DEFINITION Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 82
of 155.
ACCESSION AE005258 AE005174
VERSION AE005258.1 GI:12513751
KEYWORDS
SOURCE Escherichia coli O157:H7 EDL933.
ORGANISM Escherichia coli O157:H7 EDL933
Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE
1 (bases 1 to 13501)
Perna, N.T., Plunkett, G., III, Burland, V., Mau, B., Glasner, J.D.,
Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J., Kirkpatrick, H.A.,
Postell, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, J.,
Grobleck, E.J., Davis, N.W., Lim, A., Dimalanta, E., Potamousis, K.,
Apodaca, J., Anantharaman, T.S., Lin, J., Yen, G., Schwartz, D.C.,
Welch, R.A. and Blattner, F.R.
Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
Nature 409 (6819), 529-533 (2001)
PUBMED 11206551
MEDLINE 21074935
TITLE Direct Submission
JOURNAL Submitted (22-OCT-2000) Laboratory of Genetics, University of
Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
AUTHORS Location/Qualifiers
1. 13501
/organism="Escherichia coli O157:H7 EDL933"
/strain="EDL933"
/serotype="O157:H7"
/db_xref="taxon:155864"
/feature="enterohemorrhagic"
<1. 7576
/note="O-island #36: Region of the EDL933 chromosome not
homologous to E. coli K-12 M61655; Cryptic prophage
CP-933K; Includes one copy of the 13 bp direct repeat that
flanks the prophage"
66. 665
/feature="10mk"
/note="20981"
66. 665
/feature="10mk"
/function="putative membrane; other or unknown (Phage or
prophage Related)"
/note="Residues 1 to 199 of 199 are 72.86 pct identical to
residues 1 to 199 of 199 from Genpept 118 :
gi17532789|gb|AA63231.1|AF151091_2 (AF151091) Lom
[Prophage P-Elba]"
/codon_start=1
/transl_table=11
/product="putative outer membrane protein of prophage
CP-933K"
/protein_id="AAG55137.1"
/db_xref="GI:12513752.1"
/translation="MKRVCAIILSAICIAVSGVPAMASEHOSTLSAGILHASTARG
SDILNINVKRYETDTLGLITFSYANAEDQKTHYSDTRWHDYVNRNWFSSVAG
PSYRVNEMFSAYMAGVAYSRVSTFSGDYFRVDNKRKTHDYLTLGSDDARYSNTSLAW
GAGVQNPTEFSAVDVAVEKSGSGDWKDFIVGVCYKF"
719. 2041
/feature="20982"
719. 2041
/function="putative structure; Structural component (Phage
or Prophage Related)"

```

```

gene
CDS
66. 665
/feature="10mk"
/function="putative membrane; other or unknown (Phage or
prophage Related)"
/note="Residues 13 to 326 of 326 are 55.73 pct identical
to residues 18 to 331 of 336 from Genpept 118 :
gi16960367|gb|AA633527.1|AF170116) hypothetical protein
predicted by Glimmer [Salmonella typhimurium lr2]"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55140.1"
/db_xref="GI:12513755"
/translation="MLSPILRTFHNISVNIYVSSPCQVSPFAGKEVELKVIDEKPILF
OMFEINPERYKDEVPYNTKQHPYLDVNTAARRESRMIGIVDGEFSNOKTARS
KLERDFENVMIYLRDVEDFSMDKRLSDIYDITCEQRLTRNEDRDELILKLELR
EISKQDSLISWYAKRRNHAMDFERNLALAKAGIFPCYTNTRNHHGISEGGLYLD
ADMILITGLKGIYADGISMHYDRNRDENVNENSI1YVNRNHPALLEGLSEFMSKVD
AHYVDYGKGVKGYKVFNTPLHNYNHFCDFIEFNHPIIMNTSOTCSW"
3530. 4522
/feature="20986"
3530. 4522
/feature="20986"
/function="orf; Other or unknown (Phage or Prophage
Related)"
/note="No significant matches"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55141.1"
/db_xref="GI:12513756"
/translation="MKIPSIOSNFNSPAGSAPFIAPNRAENAYADYLDIGRIPL
SAADLSNYEESYIRAVHDSRSLIDOHVDMIGNVLDALRSOFTPRAYSVGIHNEK
VHIGCIKTRNEYLEESSVKTDIDQSLTCHNEIYKDYVGOPIPIPCAGENDNEPY
VSFSAVPDSTEMCSWOGILHEIITHIVTSSDPSGGSNIELEGTETELARVNOELG
WSVPDFKGYAEPEEREAHLRLNLALROAMKHEENEFERLGTISDRYEASDPFT

```

```

gene
    CDS
        5350..6231 /gene="20989"
        5350..6231 /gene="20989"
        /function="orf; Other or unknown (Phage or Prophage Related)"
        /note="No significant matches"
        /codon_start=1
        /transl_table=11
        /product="unknown protein encoded by prophage CP-933K"
        /protein_id="AAG55142.1"
        /db_xref="GI:12513757"
        /translation="MLSPYSVNLGCSMSNLTNRILNSPDRVLSSVRDAVAHSDGAQAV
KVGNTRYRVATDNKFCVTRESHGCPFNLLHRIGMPKGELISRKLEWLNAPSAAMAA
EKGIVYHSRNPDLPPDYDAPPPLPSVDYRIARLSPGVICKGCNAVVYPADPATYIKK
FTTOSNEEVTESEVCFENQYGACSAEKITYNGNDIITIRBDKINGESLNISSIPAL
AEHAUYDKFEDRLFOCKGLFPDTETINVLIDRAKNKEFNPTDIISTNVSDRSSEQIMQ
SYHGKKDLISRVLSTKI"
        6462..7160 /gene="%0990"
        6462..7160 /gene="%0990"
        /function="orf; Other or unknown (Phage or Prophage Related)"
        /note="No significant matches"
        /codon_start=1
        /transl_table=11
        /product="unknown protein encoded by prophage Cp-933k"
        /protein_id="AAG55143.1"
        /db_xref="GI:12513758"
        /translation="MRPTSMLYLHOSSRSSMSGDDIESLVAKSSVOVMIKNNPOLREF
OGTDHNIIQQILEADLKIGSETGRVILLNALESISRKSFEYNVIHLNSSRGVAARD
IDAENHKRTGSDFHCNALNAVPEPCGEGISVDFFATTVEFHLLHFVNLINEERLKVES
SRAEOSXKSPLLIEEARTVLGAFSEEVLSENKHEETGMPRRTSYPDSDLIHDDNTM
VSIFOOQLRHLPKL"
        complement(7668..8144)
        /gene="ybhb"
        /note="20992"
        complement(7668..8144)
        /gene="ybhb"
        /function="orf; Unknown function"
        /note="Residues 1 to 158 of 158 are 99.36 pct identical to residues 1 to 158 of 158 from Escherichia coli K-12 Strain MG1655: B0773"
        /codon_start=1
        /transl_table=11
        /product="orf, hypothetical protein"
        /protein_id="AAG55144.1"
        /db_xref="GI:12513759"
        /translation="MKLIINDIRDGDCLKPHRYHVENCGCYGDNISPILANDVIDPAIGTK
SFVYTCTDPDAPTSGSMHHWYVNLPADTRVLPQGFSGLVAMPDVGVLQRTPDGKTG
YDGAAPKRGETHRRIETVALDERIIDVDEGASGAMGFNVHFSLASAITMFMS"
        complement(8203..9492)
        /gene="bioa"
        /note="20993"
        complement(8203..9492)

CDS
Query Match      100.0%; Score 25; DB 1; Length 13501;
Best Local Similarity 100.0%; Pval No. 0.00012;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy       1 ggaatcgtctttagaacgcagcaacg 25
|||||
Db       9641 GGATCTGCTGTTTGAAGCCACAGCAg 9665

RESULT 15
LOCUS     AP002553          297816 bp            DNA             BCM           07-MAR-2001
DEFINITION   Escherichia coli O157:H7 DNA, complete genome, section 4/20.
ACCSSION    AP002553 BA000097
VERSION     AP002553.1 GI:13360211
```

KEYWORDS	EScherichia coli O157:H7 (strain:O157:H7, sub_strain:RIMD 0509552)
SOURCE	DNA.
ORGANISM	Escherichia coli O157:H7
REFERENCE	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.
AUTHORS	1 (sites) Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S., Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T., Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and Shinaagawa,H.
TITLE	Complete nucleotide sequence of the prophage VT2-Sakai carrying the verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7 derived from the Sakai outbreak
JOURNAL	Genes Genet. Syst. 74 (5), 227-239 (1999)
MEDLINE	20198780
REFERENCE	2 (sites) Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M., Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinaagawa,H. and Hayashi,T.
AUTHORS	Comparative analysis of the whole set of rRNA operons between an enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an Escherichia coli K-12 strain MG1655
TITLE	Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
JOURNAL	20557356
MEDLINE	3 (sites) Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S., Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T., Yamamoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and Shinaagawa,H.
REFERENCE	Complete nucleotide sequence of the prophage VT1-Sakai carrying the Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli O157:H7 strain derived from the Sakai outbreak
AUTHORS	Gene 258 (1-2), 127-139 (2000)
TITLE	20564182
JOURNAL	4 (sites) Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K., Yokoyama,K., Han,C.-G., Ohtsudo,E., Nakayama,K., Murata,T., Tanaka,M., Tobe,T., Iida,T., Takami,H., Honda,T., Sasakawa,C., Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and Shinaagawa,H.
MEDLINE	Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic comparison with a laboratory strain K-12
REFERENCE	DNA Res. 8 (1), 11-22 (2001)
AUTHORS	21156231 5 (bases 1 to 297816) Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shinaagawa,H. and Hayashi,T.
TITLE	Direct Submission
JOURNAL	Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome Information Research Center: 3-1, Yamadaoka, Suita, Osaka 565-0871, Japan (E-mail:ken@gen-info.osaka-u.ac.jp, URL:http://www.gen-info.osaka-u.ac.jp/, Tel:81-6-6879-8365, Fax:81-6-6879-2047)
MEDLINE	21156231
REFERENCE	genome project.
AUTHORS	Location/Qualifiers
COMMENT	1. .297816
FEATURES	/organism="Escherichia coli O157:H7"
SOURCE	/strain="O157:H7"
	/sub_strain="RIMD 0509552"
	/db_xref="taxon:83334"
	79. .1245
	/gene="ECS0753"
	79. .1245
	/gene="ECS0753"
	/note="similar to SUCC_ECOLI g111786948 percent identity 100 in 388 aa (Conserved in E.coli K-12)"
	/codon_start=1
	/transl_table=11
	/evidence=not_experimental
	/product="succinyl-CoA synthetase beta subunit"
	/protein_id="BAB34176.1"
	/db_xref="GI:13360212"


```

gene
CDS
/translation="MNLEHYOKOLFAPVGLPAPYVACTTPREAEASIKGAPW
VKQYHAGRGKAGCVKVVNSKEDIKFAENMLKRLVYTOTDANGOVYNOLIEBAT
DLAKELYLGAADVRRVRRVMASTEGVEIEKVAETPHLHKVALDPLGPMYOG
RLAKELYLEGLVOOFTKIFMGLATIFLEKVALIEINPLVITKOSDLIDGKIGA
DGNALFROPDLREMOQSOEDPREAOQAMELYNALGICMANGAGLMDGTIDIA
KLHGEPAFLDVGGAATKERYTEAFKILISDDPKAVLVNPGIVRCDLIADIIIG
AAVEGVNPPVVRLENNALGAKKLADSGNLITAAAGLIDAAQOVVAANEK"
1245..2114
/gene="ECS0754"
1245..2114
/gene="ECS0754"
/translation="similar to SUCD_ECOLI g11786949 percent identity
100 in 289 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/translation="not experimental"
/product="succinyl-CoA synthetase alpha subunit"
/protein_id="BAB34177.1"
/db_xref="GI:13360213"
/translation="MSILDKNTKVICGFTSGTSPHSEQAINGTKMVGTPFGK
GTHLGLPVNTVREVAATGATASVITVPAPFCDSILIEDAGIKILITTEGIP
LMLTVKVLDEAGVAMIGPNCVITPBECKIGIQPHIHPKGVGIVSRGLTVE
AVKQTDYGFSGTCVIGGDPIDGSNFTDILEMEKDPOTEIIVMIGEIGSABEEA
AAVKEHVTKPVVGYIAGVTAPKCKRMGAGAILAGCKTADKFALEAACVKTIRS
LADIGALKTVLK"
2781..3686
/gene="ECS0755"
2781..3686
/gene="ECS0755"
/translation="probable transcriptional regulator, similar to
transcriptional regulators e.g. glycine cleavage system
transcription activator (gcv operon activator) -
Escherichia coli g11417043|sp|P32064|GCVL_ECOLI percent
identity 31 in 300 aa"
/codon_start=1
/transl_table=11
/translation="not experimental"
/product="putative transcription regulator"
/protein_id="BAB34178.1"
/db_xref="GI:13360214"
/translation="MRGKIPKTELVTFEVVARHESYTRAEBELATQSAVPROVAL
EEFLRPLFSSHKKRIFLNDTKYVIGIVKELNKLERTNTIMTQPVQVTELVN
PFSSTHMLIPNHEPFLKLPDIIVNLSANNQDIPNREDAVIMRENCAVAEEY
LPEEETLPVSCSILAMSDOKLSVALTELLPDLHSTRTITGMEEMFALSGVSPLVN
NCRPDLISMLTAAVNSNGVALPRFALQHLDSGDWYIPCDVPIRTGNRIIMTQW
EKS DSPHLOQFREWILAKSVQDEM"
complement(3720..4322)
/gene="ECS0756"
complement(3720..4322)
/gene="ECS0756"
/translation="probable cob(1)lamina adenosyltransferase, similar
to cob(1)lamina adenosyltransferases (cortinoid
adenosyltransferases) e.g. [Escherichia coli]
g11151481|sp|P13040|BRUR_ECOLI percent identity 67 in 200
aa"
/codon_start=1
/transl_table=11
/translation="not experimental"
/product="putative cob(1)lamina adenosyltransferase"
/protein_id="BAB34179.1"
/db_xref="GI:13360215"
/translation="MEARISTEERHVOOROKIKPOVTPVAAATEKKGILIVFGNGK
KSTRAAGTVRAVGHGKTVGVAOYIKGQNDNGYINLLOPLGVEFHIMGCFMTEQNS
QADIDAKEWSESKRMLADKRYDLVLDLITVMLAYHLDLEEVYIASLONRAQOSV
IVRGCHSOLIMADTVSEIRPVKHAFMGIAQDQIDW"
complement(4332..5984)
/gene="ECS0757"
complement(4332..5984)
/gene="ECS0757"
/translation="probable fumarate hydratase, similar to fumarate
hydratases e.g. fumarate hydratase class I, aerobic
(fumarase) - Escherichia coli
g111205981|sp|P00923|FUMA_ECOLI percent identity 68 in 545

```

```

aa"
/codon_start=1
/transl_table=11
/translation="not experimental"
/product="putative fumarate hydratase"
/protein_id="BAB34180.1"
/db_xref="GI:13360216"
/translation="MSKPFIMQEPFLONKDGTEYTLISDQITVTELDGEVETIAPE
ALTLARATFEASFPLRSARLQOVAIILNDQOASSNDKYVALDILPNAEVAKVLP
NCQDTFATTVASKGQOIVTGNDNAELSKGISTFOENNLRSQNPALDMEVNTQ
TNLPAQIDISAVAGDEYHFLCVNKGGSANAKALYQETKSLDPEKTLALIKMSL
GTACPPYHIAFVYGLSADQTLKIAKIASKRYDNPSTGNEQGAFFRIIEKYL
EASQOFGIAFGKYPFAHDIRVRLPRHGSGCP IAMALSCADRNIAKINHGVL
EKLHNPGQVTPASLRENNAOHYOLDNLPVMDOLRLPGTVFSLSGPVAVAR
DIARAKTKARLDSGEPPEYIKHIIYVYAGPATRPMACGSGPTGSGMDYIDTF
QAAQGSVLMLSKGRNSQVTDACHKHGFFNGSIGGAALLAQDYKASLNCLEYPBLG
MEAVMMMEVENLPAFILLVDKGNFESQFEOQHRKASCAPAGH"
complement(6092..7372)
/gene="ECS0758"
complement(6092..7372)
/gene="ECS0758"
/translation="probable transport protein, similar to
glutamate/aspartate transport proteins (proton glutamate
symport proteins) e.g. [Bacillus stearothermophilus]
g11214671|sp|P24943|GLTP_BACST percent identity 38 in 416
aa, also similar to C4-dicarboxylate transpor"
/codon_start=1
/transl_table=11
/translation="not experimental"
/product="putative transport protein"
/protein_id="BAB34181.1"
/db_xref="GI:13360217"
/translation="MKRSITPTMILLALVYGMIIIGVYINNTASPEKTAKIYAQESIFT
TIFUKLIMITAPLVSTLVYVIGIAKKMGDAKALGIFSKTULFPCASLISRTAIGTV
NFPMPGTINFAHGAETTVAAPEFTLKVFIHAPPTSIVDMANNEILYQVEFI
FLGCSLTAIGKGSALVHALDSLAHMKILGYVLPAPLFAISALAEKGLAV
VSAGIFMEGFYFMILLVNLIGLAIYVGCIRLRALSEPALLAFPTSSSEAP
GLTELBQFGVSPKIASFVLPFGISFNLSGMAVCAATYVIAQACNIIHSIGQITM
LIIIMLSKMGAVPRASVYIAATINQFNIPGAILLIMQVDFELDMGSRATVMSN
AMGAAMVSRMGEHGECCRGKALKPNSNALP"
complement(7533..7850)
/gene="ECS0759"
complement(7533..7850)
/gene="ECS0759"
/translation="unknown"
/codon_start=1
/transl_table=11
/translation="not experimental"
/product="hypothetical protein"
/protein_id="BAB34182.1"

```

Query Match 100.0%; Score 25; DB 2; Length 297816;
Best Local Similarity 100.0%; Pred. No. 0.00012;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	99atctgcgtgttgaaagcagcagcag 25
DB	92373	GGATCTGCTGTTTGAAGCGCAGCAG 92397

RESULT 16

LOCUS	ECOBIOB	128 bp	DNA	BCF	03-JAN-1995
DEFINITION	Escherichia coli	biobin (biob) gene, early terminator region.			
ACCESSION	M27731.1	GI:341755			
VERSION	M27731.1	GI:341755			
KEYWORDS	biob gene; biotin.				
SOURCE	Escherichia coli				
ORGANISM	Escherichia coli				
REFERENCE	1	(bases 1 to 128)			
AUTHORS	Nath, S. K.				

TITLE	Attenuation of transcription of biotin genes in <i>Escherichia coli</i>
JOURNAL	Can. J. Microbiol. 34 (12), 1288-1296 (1988)
MEDLINE	89167942
FEATURES	
source	location/Qualifiers 1..128

```

/gene="b1of"
/note="7-keto-8-amino-pelargonic acid synthetase; similar
to BioF"
/codon_start=1
/transl_table=11
/product_id="7-KAPA synthetase"
/protein_id="AAC53590.1"
/db_xref="GI:12407615"
/translation="MTWQQRIDALDARRADALRVRTPEVNGAFRLIREQRICNF
SGNDYIGLSQHEPIIRAMOGABOYGVSGSGCHSVGYTHAQLEHLAGMTGYPRA
LFTSGFAANOALIALTDKEDRIADRLSHSLLEAASHSPALRRFANHYAQDS
LIDPOGQOLVYVEGIFESMDGAPLGLVHAAGQSGWLLVDADHAGLVGIVEGRG
SASQCKPELLVITFGKGVSGCAVLGSDADYLLQFARHLITSTMPAQAVALL
LAAPNTRSEDAKRRORLALIDOFRAGRDLPVITTSQSIQPLVGENARALHL
AQRIRGIVWTALIRPPYPAQTARLRLITTAHQPADITALEALHAGE"
6207..6962
/gene="b1oc"
6207..6962
/gene="b1oc"
/gene="b1oc"
/codon_start=1
/transl_table=11
/product="b1ocin biosynthesis B1oc-1like protein"
/protein_id="AAC53591.1"
/db_xref="GI:12407616"
/translation="MSQVKNQATAAAGRAAGSYSRHDALOROSADGLRALLENFA
SYLDAGCGPSNSLARMROTSVYTDMLCAPMEERARNOADSYLOADEALPLADA
QFSLAMSHLAVOXKASLPALAEIMRTQLGVAFTLLNGSLPELNQWQAVDTP
HANRFLSGQOVNDCALSGOYRCADVOTVTLFDALSMHSLKIGATHLHAGRAAPL
TRGLOQLAOWMPQOAGKFPYSYOLFGLIIEED"
6940..7680
/gene="b1od"
6940..7680
/gene="b1od"
/gene="b1od"
/codon_start=1
/transl_table=11
/product="delthiobiotin synthase B1od"
/protein_id="AAC53592.1"
/db_xref="GI:12407617"
/translation="MELSNVTERFVTGTDEYGVKTYASTALLQOANALGRHAGYRP
VAGSEMTAQLRNSDALALORNSLALSTPAVNPYFAEPTSHIVASQDRIFEA
TLASGRLALETOAQWVLVEGAGWFTPLSQEPFADWMAERLPIVLLVGVKLCIHH
AMITAOAIRQAGLHFAGWANGVYVPGKRHAHEYMATLRYLSAPLLGEIPWLGROADT
SAVGQYDLTALTPSAPSSGSGAGHPAR"
complement(7619..>8227)
/gene="elsa"
complement(7619..>8227)
/gene="elsa"
/note="putative urea/short-chain amide ABC transporter"
/codon_start=1
/transl_table=11
/product="ABC transporter-1like protein"
/protein_id="AAC53594.1"
/db_xref="GI:12407619"
/translation="DCXLRTRXGENXIVSCITAGLLPIASGTMMWHEAPGAVNAGTEI
XYVPDRIEFSQLEVENTLIIAOMAKGIAGGSLSRDVEEPPELPIYRLOLRASGLND
QOYOLALANAIVTRPRLLIIDEPSRGCTGLTLYKALLRFNRELIGLIVLADQHS
LIRVADREFCLIRGRNVAQGRVSELDPLIAHMGPIASAO"
15 others
ORIGIN
1512 a 2642 c 2605 g 1453 t

```

Query Match 84.0%; Score 21; DB 2; Length 8227;
 Best Local Similarity 100.0%; Pred. No. 0.031;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 ctgctgttgaagcgagcag 25
 ||||||||||||||||
 Db 4095 CTGCTGTTTGAAGCCGACG 4115

RESULT 18
 EMBIOTOP2
 LOCUS EMBIOTOP2 1041 bp DNA BCT 04-NOV-1996

```

DEFINITION Erwinia herbicola biotin synthetase (b1ob) gene, complete cds.
ACCESSION U38648
VERSION U38648.1 GI:1228112
KEYWORDS 2 of 2
SEGMENT Erwinia herbicola strain=Eh10.
SOURCE Pantoea agglomerans
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Pantoea.
REFERENCE 1 (bases 1 to 1041)
AUTHORS Wu,C.H., Chen,H.Y. and Shuan,D.
TITLE Isolation and characterization of the Erwinia herbicola bio operon
and the sequences of the b1oa and b1ob genes
JOURNAL Gene 174 (2), 251-258 (1996)
MEDLINE 97045821
REFERENCE 2 (bases 1 to 1041)
AUTHORS Shuan,D.
TITLE Direct Submission
JOURNAL Submitted (14-OCT-1995) David Shuan, Biology, Natl. Sun Yat-Sen
University, Kaohsiung, Taiwan, ROC
FEATURES
Location/Qualifiers
1..1041
/organism="Pantoea agglomerans"
/strain="Eh10"
/db_xref="taxon:549"
1..1041
/gene="b1ob"
1..1041
/gene="b1ob"
/codon_start=1
/transl_table=11
/product="biotin synthetase"
/protein_id="AAC44534.1"
/db_xref="GI:1228114"
/translation="MAHHRKWTLSOYTELFKPLDILFEAQTIRHQHPDROVST
LSTKTKGAPEDCKYCPQSSRYKTGLETERLMEVDQDLSAKRANAGSTRFCMAAM
KNHERDMYPLEQWVGVKAMGLEACMTLGTIDETQOQLASAGLDYNNHNDTSPEF
VGNILITRYQERLDTLDKRVADGIVCGGIVGAEYKFKDRAGLLQLANLPTPES
VPNMLVKKYKGPRLDNDVDADFRTIAVARIMMPTSFVRLSAGREOMNQTQAMC
PAGANSTVGGKLLTTPREDKDVQFLRKGLNPGQVDVTGNEQOKLEQIFN
ADDQRYNNAAL"
BASE COUNT 230 a 322 c 308 g 181 t
ORIGIN

```

Query Match 72.0%; Score 18; DB 3; Length 1041;
 Best Local Similarity 100.0%; Pred. No. 2;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 ctgctgttgaagcgagcag 22
 ||||||||||||||||
 Db 67 CTGCTGTTTGAAGCCGACG 84

RESULT 19
 LOCUS E00893 1121 bp DNA PAT 29-SEP-1997
 DEFINITION Genomic DNA encoding biotin Synthetase.
 ACCESSION E00893
 VERSION E00893.1 GI:2169154
 KEYWORDS JP 1986149091-A/1.
 SOURCE Escherichia coli.
 ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Escherichia.
 REFERENCE 1 (bases 1 to 1121)
 AUTHORS Hirano,Y., Kojima,T. and Kimura,H.
 TITLE DUPLICATION OF BIOTIN SYNTHASE, BACTERIUM CONTAINING SAME AND
 PRODUCTION OF BIOTIN
 JOURNAL Patent: JP 1986149091-A 1 07-JUL-1986;
 NIPPON SODA CO. LTD
 OS Escherichia coli
 PN JP 1986149091-A/1

PD 07-JUL-1986
 PF 24-DEC-1984 JP 1984272605
 PI HIROMO YOSHIIKO, KOJIMA TAKAKAZU, KIMURA HITOSHI PC
 C12M15/00, C12N1/20, C12P13/18, (C12N1/20, C12N1/19), (C12P13/18, PC
 C12R1/19)
 CC strandedness: Double;
 CC topology: Linear;
 CC hypothetical: No;
 CC anti-sense: No;
 CC *source: strain=Escherichia coli Nsl01;
 CC feature is identified by experimental;
 FH Key Location/Qualifiers
 FT
 CDS 42..1079
 /product="biotin synthetase".
 Location/Qualifiers
 1..1121
 /organism="Escherichia coli"
 /db_xref="taxon:562"

BASE COUNT 289 a 296 c 325 g 211 t
 ORIGIN

Query Match 72.0%; Score 18; DB 10; Length 1121;
 Best Local Similarity 100.0%; Pred. No. 2;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gagatcgtgttgaagc 18
 |||||||
 Db 104 GGATCTGCTGTGAGC 121

RESULT 20
 LOCUS AF281080 1614 bp mRNA VRT 01-OCT-2000
 DEFINITION Xenopus laevis homeobox transcription factor Nkx2-1 (Nkx2-1) mRNA,
 complete cds.
 ACCESSION AF281080
 VERSION AF281080.1 GI:10442649
 KEYWORDS
 SOURCE African clawed frog.
 ORGANISM
 Xenopus laevis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 Xenopodinae; Xenopus.
 1 (bases 1 to 1614)
 Small, E.M., Vokes, S.A., Garlick, R.J., Li, D. and Krieg, P.A.
 Developmental expression of the xenopus Nkx2-1 and Nkx2-4 genes
 Mech. Dev. 96 (2), 259-262 (2000)
 2 (bases 1 to 1614)
 Small, E.M., Vokes, S.A. and Krieg, P.A.
 Direct Submission
 Submitted (21-JUN-2000) Cell Biology and Anatomy, University of
 Arizona Health Sciences Center, 1501 N. Campbell Ave., Tucson, AZ
 85724, USA

FEATURES
 source
 1..1614
 /organism="Xenopus laevis"
 /db_xref="taxon:8355"
 /note="expressed in ventral forebrain, lungbud primordia
 and thyroid anlage"
 1..1614
 /gene="Nkx2-1"
 162..1205
 /gene="Nkx2-1"
 /note="thyroid transcription factor; TTF-1; T/ebp; chicken
 and mouse Nkx2-1 ortholog; contains homeodomain"
 /codon_start=1
 /product="homeobox transcription factor Nkx2-1"
 /protein_id="AA017405.1"
 /db_xref="GI:10442650"
 /translation="MSMSPKHTTPSVSDILSPFESYKKVAMEGAGIGAPLTAAAYR

OSVSAQSMQOQMGNGCPVSAHYMTAAGVQPOLSHTTMGYCGNGLNGLMSLEPP
 YDUTMRKNSVATGAYGALNPPRSTISRFMGPGGMMGLGMMGLGVMGSMPLQA
 TPRKRRVLESOAYVLERRRFRKQKQYASAPREHLSMTILPTPOVLTWQNHRYKM
 KRQAKRASQOOIOODNSCCOQOQSPRYAVPVLYKDGKPCQAGSNPTAALQSHQOQ
 TATATITTINGLGPQHSHQSHQNSAGSPDLVPHSNPSSLQNGVTSLSHLNSSSDYGS
 AMSCSTLYGRTM"

BASE COUNT 475 a 407 c 399 g 333 t
 ORIGIN

Query Match 68.0%; Score 17; DB 8; Length 1614;
 Best Local Similarity 100.0%; Pred. No. 8.1;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 tcgcgtgttgaagc 20
 |||||||
 Db 717 TCTGCTGTTGAGCGC 701

RESULT 21
 LOCUS AE001220 14268 bp DNA BCT 16-JUL-1998
 DEFINITION Treponema pallidum section 36 of 87 of the complete genome.
 ACCESSION AE001220 AE000520
 VERSION AE001220.1 GI:3322705
 KEYWORDS
 SOURCE Treponema pallidum.
 ORGANISM Treponema pallidum.
 Bacteria; Spirochaetales; Spirochaetaceae; Treponema.

REFERENCE 1 (bases 1 to 14268)
 Fraser, C.M., Norris, S.J., Weinstock, G.M., White, O., Sutton, G.G.,
 Dodson, R., Gwin, M., Hickey, E.K., Clayton, R., Ketchum, K.A.,
 Sodergren, E., Hardham, D.M., McLeod, M.P., Salzberg, S., Peterson, J.,
 Khailak, H., Richardson, D., Howell, J.K., Chidambaram, M.,
 Uitterback, T., McDonald, L., Artlich, P., Bowman, C., Cotton, M.D.,
 Weidman, J., Smith, H.O. and Venter, J.C.
 Venter, J.C. et al.
 Complete genome sequence of Treponema pallidum, the syphilis
 spirochete
 Science 281 (5375), 375-388 (1998)
 JOURNAL 98332770
 MEDLINE
 REFERENCE 2 (bases 1 to 14268)
 Fraser, C.M., Norris, S.J., Weinstock, G.M., White, O., Sutton, G.G.,
 Dodson, R., Gwin, M., Hickey, E.K., Clayton, R., Ketchum, K.A.,
 Sodergren, E., Hardham, D.M., McLeod, M.P., Salzberg, S., Peterson, J.,
 Khailak, H., Richardson, D., Howell, J.K., Chidambaram, M.,
 Uitterback, T., McDonald, L., Artlich, P., Bowman, C., Cotton, M.D.,
 Weidman, J., Smith, H.O. and Venter, J.C.
 Direct Submission
 Submitted (06-MAR-1998) The Institute for Genomic Research, 9712
 Medical Center Dr, Rockville, MD 20850, USA

FEATURES
 source
 1..14268
 /organism="Treponema pallidum"
 /db_xref="taxon:160"
 98..2149
 /gene="TP0421"
 98..2149
 /gene="TP0421"
 /note="similar to GB:AE000783 percent identity: 27.26;
 identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=1
 /product="conserved hypothetical protein"
 /protein_id="AAC65409.1"
 /db_xref="GI:3322706"
 /translation="MRGWCISRSYVACAGALLLFAVGAQEHAPPVRYDTALAGEF
 RGVLYAYRGAFNDAIQLFERALSAAFRNLLILEMIGNAYRSGIAGALHONGARD
 LGYGMALRNKIEVVOORDPAPDSADALHIFESSESFHAVERGTGLFRRLTSLCALD
 GTGWSAYVGSNELIREPDVNGRVYARTRGPEGRPDVYOTSGGLVSEPSADSDIC
 RLTKREGRLSPGKSGKGVQOLIGPOFLADRIDNIVYDFGNARVAIVPDGAPLPH
 FGKSAHFPEFSAPGAIYIDELVAYDAKGAIVYVDGAGNVYRVLVAEGTLKHVES
 VRAWNGRLVSLPREVYVVDVGLASLHTIARIGNAPIRLTAAVPDANGVLLADYKNE

RLEIYSRISLGLGLFVHFERVHADRPSPVNVSVHRECPVVGGLDVSFLLTEEH
 RPVSEMRLOGAAYLNDTCSSILYDRSPATQERMLYTRVIGELAQAGEGTSLSVS
 AAOSPVELEGTFLAOLQAOPLRLAKSSSQMFEDGLGLAGTLVADYKRAVFLSL
 GALTQDPAOYGVQDLAAVLTNNETSPVIOLOFTVPLVELGYLVOKTRGAVSVYAO
 EELGALVAHVHSGAAGTYALSRLPTDFGRAVMPVEAEVRLSRSGRDEGTGYEAPL
 E
 2133. 3158
 /gene="TP0422"
 2133. 3158
 /gene="TP0422"
 /note="hypothetical protein; identified by Glimmer;
 putative"
 /codon_start=1
 /transl_table=11
 /product="T. pallidum predicted coding region TP0422"
 /protein_id="AAC65418.1"
 /db_xref="GI:3322715"
 /translation="MRRLSDRLRGMTGKGSMAVRLCYLSTQVGLAEPALRLAGVW
 ENARPIVAPTPADAMRFSIMLVHPPYAVYERVAVSLSLSDARADVLPPPAOK
 AERVLVLSFEGRVVQPLVLCRMGDALFCSPFQVSYGTRSGSYSSFOITRYDDPL
 FEFWIEGSAELRAVYSSOAGELNAFFETANFYRFRWRDDALAFPAORAYFTADD
 GVTYEPVOYVGAAYVSCITGSRVVRNFOGTVEVRSIDGSKRLMLRRGAGPGS
 YAVGAVYPRQGFDESALALRYDGOFLAIGEPFLKRSVAVHDVTAFLIDAHMKRAPRA
 PLVPDF"
 3183. 3989
 /gene="TP0423"
 3183. 3989
 /gene="TP0423"
 /note="similar to GB:AE000783 percent identity: 39.66;
 identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AAC65410.1"
 /db_xref="GI:3322707"
 /translation="MRVFSWMKOKDGERERAGALOGACIGTPPEHROFALVGRSGS
 GSFRAQIVMRFRDIAFVVDGLLREDDIINGRSKORITLLAIRVALFALVGRSGS
 VVARVITYFLCSQNKVLLITSEMYRIALRVGLPERGHITRIEDIASSEELACA
 LTRHVEHVIIPVPSGEVRSKYPKIFTERIKLRLRREGAERIGMAHALHHEGLKR
 ACSGAHVAHEKSIVRPPSCNLRVVAATEGQDASPVVPPHEELAQO"
 4192. 4890
 /gene="TP0424"
 4192. 4890
 /gene="TP0424"
 /note="similar to GP:1510115 percent identity: 37.59;
 identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="V-type ATPase, subunit E, putative"
 /protein_id="AAC65411.1"
 /db_xref="GI:3322708"
 /translation="MLGESREAEARTVRAAREAEARTVRAAREAEERLESSIALSQ
 ASRNVLSFQDSVTRSLRAITSMETQAQADAGLRLIPIVVSAMVQAGEGDELEILS
 PALRLLEGVFCALGEOISAGVELRSDCLAGFRIVPAEGSYDFFSAAVAOLFS
 SYVASARVAEVLILRRSCDVFLILPYNATISSMRGASFALSGFGLCATPLGSAAGC
 CGMKTLVVSRAQA"
 4868. 5128
 /gene="TP0425"
 4868. 5128
 /gene="TP0425"
 /note="hypothetical protein; identified by Glimmer;
 putative"
 /codon_start=1
 /transl_table=11
 /product="T. pallidum predicted coding region TP0425"
 /protein_id="AAC65419.1"
 /db_xref="GI:3322716"
 /translation="MCPVPARDSTGSRVLDRMGEFFYALRCALARARARLRMDPTVE
 VROCVAQESVAVARAVSLDDPLRAGVFFERRALFLRCGV"
 5270. 7039
 /gene="TP0426"
 5270. 7039
 /gene="TP0426"
 /gene="TP0426"

/note="similar to GB:AE000783 percent identity: 63.40;
 identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="V-type ATPase, subunit A (atpa-1)"
 /protein_id="AAC65412.1"
 /db_xref="GI:3322709"
 /translation="MTQKGISAVNGNMVSVTEEGVSLNEGYVHVCNARLKEI
 RYGRERAOLOVFEITRGSVGDVEPEFDGLISVEIGGLIGYDGLDNLPLLAERK
 GYFLERGVLPALSRISSEMTTPHVSVEGEVYKGDVIGTIPBEGLKHRIIVPHMDS
 YEVEFIQTACTYRHVDIAVRDAQGHEHETLMAFRPVRPVHCTAERLAKPEPLVY
 SIPTIDFFPVAKGTCICIPGAGTVDLOHSTSRADVDVYIAACGRAGEVET
 LREPDLDIPRTGSLMERTVIVCNTSMVAERSASVGTGVLAEYRQMGDLVLL
 ADSTRMAOALREMSGRLEIPGEAPALYESIAFYRAGVRLRSGESVYIG
 CTVSPAGCNPEEPYQATLVYVAGFGLSRERSDARVPAVHPLDSKXYPSYDARA
 VAGRSFLRGAEEVOMRVYGEBSGTSMEPLVYLKSGFLDSYILOONSDEYDASVP
 VAKRHCYAIYMRVLSVLAPESKDDARAFSKUHGFIIDYNCAMNSEAFEREKEI
 RAFLQGSTKIDSEAGCIIIRGME"
 7043. 8335
 /gene="TP0427"
 7043. 8335
 /gene="TP0427"
 /note="similar to GB:AE000783 percent identity: 72.96;
 identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="V-type ATPase, subunit B (atpb-1)"
 /protein_id="AAC65413.1"
 /db_xref="GI:3322710"
 /translation="MYRYEKIESINGSVITVRAGVHYGELARVKTAFGDSIAEVNK
 LEGDVLSQVPAFGRGISTGSEVFLRGMLVNSDYLLRPFNGSGVPDGPALKE
 GRVLEGPVSNVPAKRVIAARMIRPTGIMIDVPLVYSOKPLPSSGSEYNEELARI
 AMQAEVYIVYIGGCIKHDVLYFKSALFEAGLSRAVLVYHNTASDPTVGLVWPMG
 LVAEQRALKRVDVLLTDMTNFADAKELAIIOEVPYSRQPGDLYSOLASRYK
 LAVDDAGSVTILAVTTMPGDVTHVPDNTGITIIEOFLTKNRITPESLSIKN
 VNGTRADHRLMDNMILKYAAYDALEKRSMGFMSMEDEKLLTYGGLFEADLMDS
 VNPLEEAFDGMKILGRCPPEETGIRSLINRYMWA"
 8347. 8967
 /gene="TP0428"
 8347. 8967
 /gene="TP0428"
 /note="similar to GB:AE000783 percent identity: 45.69;
 identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="V-type ATPase, subunit D (atpd-1)"
 /protein_id="AAC65414.1"
 /db_xref="GI:3322711"
 /translation="MAVRLTKNELRKOKETLTFRRPLTLQKKOOLYAEIRAVEAR
 ASEVSRADALRODYOQWAVVESGVEDARLLHVOEKKTYASIAGRIPCFLVHR
 ACPPDYLRVPLMDPTAVRKMOEVLISDLEAHVILDBOALLGAEIRTTQTVNLFERY
 KIPETRACIRKITYILGDOQVAAVYRGMSKKNLVDSRMQEDTR"
 8964. 10832
 /gene="TP0429"
 8964. 10832
 /gene="TP0429"
 /note="similar to GB:L77117 SP-Q57675 PID:1590960 percent
 identity: 25.19; identified by sequence similarity;
 putative"
 /codon_start=1
 /transl_table=11
 /product="V-type ATPase, subunit I (atpi-1)"
 /protein_id="AAC65415.1"
 /db_xref="GI:3322712"
 /translation="MIVPKVYTLVLVGSDEBSLQALRSFGAVHVLRECAEQALAE
 LHAIDARCVQALVTAQOTKNTVNGEGRVAGOVAEAEVQIVRTHSDRVELAOK
 IAOCIALHERCEPMGDPADYRALAQGHILIPVELESERYRCLPELOTLCLARIG
 GIVRCVLAQKRGGLPSPADYRALALDVSYPADYFLVRLQRECATLQRLRLAAS
 YOGARLAIROKTAIDIERFERVHLISWVSVDVOMREGTTLIAHSGYLPVSRVAFS
 BEARKEANAACCVDPMEDPVPQOIRNNRNWNLISPLMNFGLTGVGVEVVISGPFIL
 FGVFFSTIFADAGVAVTLVSGGITYLSKRKAHVSFAMCIGLYLCTLMNCAIY
 CMMPGVPOVYVASLARIAVWEISGFADAORNNQMHVCFGLGLHLCGLHLYVR
 TFRSLRVLAEFGSLIMLGMYVYVNLIVDEREYPLTGMIVGSIAGVLMFIVNRY

USVRSVADSKMKNVINALGIVNPAVMSYIRLMAVAGLCAISATVNTKTHPLFAN
FLATGIVLILFGHGLNVTMSIISVIVHGVNLTLERSNHVGLMTSIRYPPFRER"

Query Match 68.0%; Score 17; DB 1; Length 14268;
Best Local Similarity 100.0%; Pred. No. 8.3;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 gctgttgaagcgacg 23
|||||

DB 8188 GCTGTTGAAGCGACG 8204

RESULT 22
AX064067 924 bp DNA PAT 24-JAN-2001
LOCUS Sequence 349 from Patent WO0100843.
ACCESSION AX064067
VERSION AX064067.1 GI:12541779
KEYWORDS
SOURCE
ORGANISM
Corynebacterium glutamicum.
Corynebacterium glutamicum
Bacteria: Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Corynebacteriaceae;
Corynebacterium
1 (bases 1 to 924)
Pompejus, M., Kroege, B., Schroeder, H., Zeider, O. and Haberhauser, G.
corynebacterium glutamicum genes encoding metabolic pathway
proteins
Patent: WO 0100843-A 349 04-JAN-2001;
JOURNAL BASF AKTIENGESELLSCHAFT (DE)
FEATURES
Source
Location/Qualifiers
1..924
/organism="Corynebacterium glutamicum"
/db_xref="taxon:1718"
101..904
/note="RXA01757"
/codon_start=1
/transl_table=1
/protein_id="CAC25274.1"
/db_xref="GI:12541780"
/translation="MPTASPIYDVVVVVGAGISGLITALLDRAGLIKCEPACSRVGG
RAVSVOOSDFRFGATWFWLNPVLYOQLVNNIGLSTFQALFGLDELFEFLVYAPRL
KGNPIDASGRFQAGASSLALGLAOLIKRVELSLGPVHSLSEDEIYVKSQIVR
AKHVITAVPPALAEILGFTLIDPADVRAAHPOHILAVMMMAEKETLPTQAAASAGF
GHELFQOPLGHRIHMASTEVEAFGCHGEGAVRAGIOALQIGFNLKS"

BASE COUNT 214 a 252 c 239 g 219 t
ORIGIN

Query Match 64.0%; Score 16; DB 9; Length 924;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 ctgttgaagcgacg 23
|||||

DB 884 CTGTTGAAGCGACG 869

RESULT 23
YL1012084 6887 bp DNA PLN 23-SEP-1999
LOCUS Yarrowia lipolytica pex10 gene for peroxin 10.
DEFINITION AJ012084
VERSION AJ012084.1 GI:5921565
KEYWORDS peroxin 10; pex10 gene.
SOURCE Yarrowia lipolytica.
ORGANISM Yarrowia lipolytica
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Dipodascaceae; Yarrowia.
1 (bases 1 to 6887)
REFERENCE Le Bail, M. T., Gallard, C. and Nicaud, J. M. J.
Pex10 is essential for peroxisome biogenesis in Yarrowia lipolytica
JOURNAL Unpublished

REFERENCE 2 (bases 1 to 6887)
AUTHORS Nicaud, J. M.
TITLE Direct Submission
JOURNAL Submitted (20-SEP-1999) Nicaud J. M., Laboratoire de Genetique des
Microorganismes, INRA-CNRS, BP 01, Tiverval-Grignon, F78850, FRANCE
FEATURES
Source
Location/Qualifiers
1..6887
/organism="Yarrowia lipolytica"
/strain="W29"
/db_xref="taxon:4952"
complement(2293..3357)
/gene="pex10"
/gene="pex10"
complement(2293..3357)
/function="essential for peroxisome biogenesis"
/codon_start=1
/evidence="experimental"
/product="peroxin 10"
/protein_id="CAB56484.1"
/db_xref="GI:5921566"
/translation="MSDNTTIKKPIRKPPIRTERLPYGAAEIIRANOKHYESVLE
QHLYVFLKMKGVREIHQYKEELFPAKFAVYIGLCTIGVSKTIGREYTNLMYTRDRT
ALPGVVRFGVLSNTLPPYLFVRVYMGKLRKIMREYHILVEYDEPVSPEYWKER
VITFVKNPKDFKTALEGFATILAIPEYKGSYOLSKIRMGRIYFGRDKNPRIG
YMLGILIFARPAITSFYOTGRYDALILEKSYEKAGKEDEKAVVKKKSITPLE
DTGETEDKIDLEDPRQLKFIPEASRACITLCLYSTISAPACTPCGHFCWDCISBVR
KECPICRQGVQNEMLPIR"

BASE COUNT 1793 a 1841 c 1485 g 1768 t
ORIGIN

Query Match 64.0%; Score 16; DB 15; Length 6887;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 gactgctgttgaag 17
|||||

DB 4051 GACTGCTGTTGAAG 4066

RESULT 24
AE002054 11823 bp DNA BCT 22-NOV-1999
LOCUS Deinococcus radiodurans R1 section 191 of 229 of the complete
DEFINITION chromosome 1
VERSION AE002054 AE000513
KEYWORDS
SOURCE
ORGANISM
Deinococcus radiodurans.
Deinococcus radiodurans
Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
1 (bases 1 to 11823)
White, O., Eisen, J. A., Heidelberg, J. F., Hickey, E. K., Peterson, J. D.,
Dodson, R. J., Haft, D. H., Gwinn, M. L., Nelson, W. C., Richardson, D. L.,
Moffat, K. S., Qin, H., Jiang, L., Pamphile, W., Crosby, M., Shen, M.,
Vamathevan, J. J., Lam, P., McDonald, L., Utterback, T., Zalewski, C.,
Makarova, K. S., Aravind, L., Daly, M. J., Fraser, C. M. et al.
Genome sequence of the radioresistant bacterium Deinococcus
radiodurans R1
Science 286 (5444), 1571-1577 (1999)
2 (bases 1 to 11823)
White, O., Eisen, J. A., Heidelberg, J. F., Hickey, E. K., Peterson, J. D.,
Dodson, R. J., Haft, D. H., Gwinn, M. L., Nelson, W. C., Richardson, D. L.,
Moffat, K. S., Qin, H., Jiang, L., Pamphile, W., Crosby, M., Shen, M.,
Vamathevan, J. J., Lam, P., McDonald, L., Utterback, T., Zalewski, C.,
Makarova, K. S., Aravind, L., Daly, M. J., Minton, K. W.,
Fleischmann, R. D., Ketchum, K. A., Nelson, K. E., Salzberg, S.,
Smith, H. O., Venter, J. C. and Fraser, C. M.
Direct Submission
Submitted (08-NOV-1999) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
Location/Qualifiers

```
source
1. 11823
/organism="Deinococcus radiodurans"
/strain="R1"
/db_xref="taxon:1299"
/chromosome="1"
103..390
/gene="DR2207"
103..390
/gene="DR2207"
/translation="MDMSHDTFEARALQAYEMTANHTYEGKVGOGHNEARADIRRL
PILCGHDEPDHAHESKKAQRKAKAEKRRKKFADLPVAEEALGRS"
complement(449..829)
/gene="DR2208"
complement(449..829)
/gene="DR2208"
/translation="MRALFTCLYADLSAAEFYSGVLGTLHSHKVGARHLFYRLDGS
MLIIFPASAQKGDVPSHOGKPGHACLAIPREQAEWETRLRGHLDVTRYAMGER
GESLYDDPAGNVELAPASIMCL"
complement(993..1904)
/gene="DR2209"
complement(993..1904)
/gene="DR2209"
/translation="MSAHTDDAGNDSGDHSDWDADPGOKLRKTRKPOGRRLA
DITGETDIDADTIRLLDLGHTEYVCKSGKATATYATYVGRPALIKYRFA
ARSFKNDATYREGQVVDODARARMSRSKSGLEMLQFMAEYTAHLQWQAGLSV
PEPLVGPHTVATVAVLMLRLIGTAESPAPRLSPVRLTPEARARAOQSVQGMADL
LRLYAHGDSYTNLWMENTVTIIDPQTLTRANPFEKELLRDAQSLATSPFKHGL
QETAEGTLREVQRALGPAPERVLVP"
2390..2773
/gene="DR2210"
2390..2773
/gene="DR2210"
/translation="MKGYCSGLLLALLALSPLOSRAOQTSCTNYPVGFDOYCVNK
VTLQADIENTAVOKLRLKPLVSAOQTLROTGRAMMAKRUNDCVDGTYVYVNCNTHMT
VSRNLNLDRLRRCNMAAGCQPSKLR"
2813..3505
/gene="DR2211"
2813..3505
/gene="DR2211"
/translation="GB:L10328 SP:P19396 GB:M24503 PID:147866
PID:290501 percent identity: 69.30; identified by sequence
similarity: putative"
/codon_start=1
/transl_table=11
/product="rRNA methylase Spou"
/protein_id="AA11758.1"
/db_xref="GI:6460013"
/gene="DR2212"
3516..4277
/gene="DR2212"
3516..4277
/gene="DR2212"
/translation="MEANTDLARSAIEKQKLGILYRQVGGRRANAGRLNHNRA
HDULTFIDGSLPVTODVALEGRESALIASVGLALPDQADDEALTRIVYDDILT
DAEVARLPLVLDLETLIDVRADGWAAYAGSDGEABQADSRQAGRRVQPADLIAYPG
SLAERLREATLRAHFVOTHPERLIRAPGVPOFRLGLVLSVTSGRPAVPDPEGARHL
DAGOLARVALLANHMAQAQEAIALDAEAFQRTLR"
4297..5313
/gene="DR2213"
4297..5313
/gene="DR2213"
/translation="MSHGRFPRVLTNFSVTSNAQIALDESPYEYEWASHSDPHNM
LAAPHITTEGRGLGDDYVAMILGTCHEGFDIVLPCKEERELAAVRDFAAGTOL
VVPADATQRIHLEKDEFLRGDDTDILPFWPTTQSAEERDACAOLROGVYKJCK
PARGIYASGFVLTLPDLPLKGEILQULSVAAARELLAGELPVMLHTLEGAEK
SIDCEYMEGRILIRAVYRRKGAHGOYIEDRPLVANAARIQAHVALSGIFNFQTDQGS
AAHMEIINARASGLRYSMAGVNFARILLDAATGRILMDALPAVEGTGLTVAEBKGVK
AMRA"
5301..6416
/gene="DR2214"
5301..6416
/gene="DR2214"
/translation="MAGMKGVSAOELTVSLPSGTLTLOVERTLYPAPBLDYAVR
NPKRGFLPYSRVLGKHPAPATARTWRDLAASLPPLAPHEIGLATATAGLGVY
RAMROOTGOATFQHTTRKYVDRPVLLAPDEPHSRAHLLYDEPGARARRELIVD
DIVSTGTLLENLALMLARHPHTRVVLVSLTDKCPREBALCALAGYVVDVSLRGR
FAETPAPDWCPALPAVYSGADKATALLPRAARAGDVDPPLTFGLSPYDRALVIG
TGEQFPAYALAMLAPOVAECRNNAATRSPLTPGRLAHLASFNDVGDMPNTLYN
VDPAEYTRIFVAYEGACVDPALLRQLRPHAOAVRL"
6413..7195
/gene="DR2215"
6413..7195
/gene="DR2215"
/translation="MSHGRAPILAFADIDTILFOTLKLPGHGRALTPATVTRGE
AHSYCPAQATLIDLITAGAVTVIPVYTGDRPAMRRVLLPFRSMVVLHGLTILPAG
ELDSEWAAVOVAHILPILQGLTALPADMAHNDALGCRITVRADDLPEFMVYIKHPDA
DPEALALOTRMEAAVDDDTGTHPLVIANANNVSLPAPRGAQAVRMLRBRHPPGA
VLTILGDSVSDLPFLRECDFAVTRAQGLLRTAGVRLPQR"
7253..8008
/gene="DR2216"
/translation="MAPPTERYAKIRRVLSKROPTTLVLNDEVNKPHNLSAIRTCDA
VGPEAHVAPPKHGTILASFECHTYEAGSAAHKWVHPADVAASVARELOAGFQVL
ATHLSORSVYRPPDYTKPTCVLLGAKMKVSPGAAPADANVIVIPMGVOSLANSY
AAATITFEAQORLAAAGYEPQSLPELRLRSLSEWGYPLDLAGYIRREGATYALGEE
GELREPVSRPE"
3516..4277
/gene="DR2217"
3516..4277
/gene="DR2217"
/translation="MSHGRAPILAFADIDTILFOTLKLPGHGRALTPATVTRGE
AHSYCPAQATLIDLITAGAVTVIPVYTGDRPAMRRVLLPFRSMVVLHGLTILPAG
ELDSEWAAVOVAHILPILQGLTALPADMAHNDALGCRITVRADDLPEFMVYIKHPDA
DPEALALOTRMEAAVDDDTGTHPLVIANANNVSLPAPRGAQAVRMLRBRHPPGA
VLTILGDSVSDLPFLRECDFAVTRAQGLLRTAGVRLPQR"
7253..8008
/gene="DR2216"
/translation="MAPPTERYAKIRRVLSKROPTTLVLNDEVNKPHNLSAIRTCDA
VGPEAHVAPPKHGTILASFECHTYEAGSAAHKWVHPADVAASVARELOAGFQVL
ATHLSORSVYRPPDYTKPTCVLLGAKMKVSPGAAPADANVIVIPMGVOSLANSY
AAATITFEAQORLAAAGYEPQSLPELRLRSLSEWGYPLDLAGYIRREGATYALGEE
GELREPVSRPE"
3516..4277
/gene="DR2217"
3516..4277
/gene="DR2217"
/translation="MSHGRAPILAFADIDTILFOTLKLPGHGRALTPATVTRGE
AHSYCPAQATLIDLITAGAVTVIPVYTGDRPAMRRVLLPFRSMVVLHGLTILPAG
ELDSEWAAVOVAHILPILQGLTALPADMAHNDALGCRITVRADDLPEFMVYIKHPDA
DPEALALOTRMEAAVDDDTGTHPLVIANANNVSLPAPRGAQAVRMLRBRHPPGA
VLTILGDSVSDLPFLRECDFAVTRAQGLLRTAGVRLPQR"
7253..8008
/gene="DR2216"
```

CDS

7253..8008
/gene="DR2216"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAPI1766.1"
/db_xref="GI:6460021"

translation="MOLVQDRHGHTLRTHALVHTFSGADGIRETREMGRFTV
EALISGGVLEPAQFYSVGOVASEQOQGGFLRAIATAGTESAPATHLSHG
RWTEPRRHFIIEATGESGDMILDDRAFYLAODMOLGTHHSGSALSGGM
QPKLSAGLFEVSEVSEVOYIELDRTVVDGDMIMYSASLRVELRLVRCIR
NAARSGEIVFEVSGKGVFLPFLHQLSASV"

9265..10326
/gene="DR2218"
/db_xref="GI:6460021"
/gene="DR2218"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAPI1767.1"
/db_xref="GI:6460022"

Query Match 64.0%; Score 16; DB 1; Length 11823;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 ctgtttgaagcgcagc 23
|||||
Db 3308 CTGTTGAGCGCAGC 3323

RESULT 25
AE004026
LOCUS AE004026 13856 bp DNA BCT 06-JUL-2000
DEFINITION Xylella fastidiosa, section 172 of 229 of the complete genome.
ACCESSION AE004026
VERSION AE004026.1 GI:9107228
KEYWORDS
SOURCE
ORGANISM Xylella fastidiosa.
Xylella fastidiosa
Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
Xylella.
1 (bases 1 to 13856)
Silvestri,M.L., Siqueira,W.J., de Souza,A.A., de Souza,A.P.,
Terenzi,M.F., Truffi,D., Tsai,S.M., Tsunako,M.H., Vallada,H., Van
Sluys,M.A., Verjovski-Almeida,S., Vettore,A.L., Zago,M.A., Zatz,M.,
Weidens,J., and Setubal,J.C.
The genome sequence of the plant pathogen Xylella fastidiosa. The
Xylella fastidiosa Consortium of the Organization for Nucleotide
Sequencing and Analysis, Sao Paulo, Brazil
Nature 406 (6792), 151-157 (2000)
20365717
2 (bases 1 to 13856)
Simpson,A.J.G., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M.,
Alvarenga,R., Alves,L.M.C., Araya,J.E., Baia,G.S., Baptista,C.S.,
Barros,M.H., Bonaccorsi,E.D., Bordin,S., Bove,J.M., Brites,M.R.S.,
Buono,M.R.P., Camargo,A.A., Camargo,L.E.A., Carraro,D.M.,
Carter,H., Colauto,N.B., Colombo,C., Costa,F.F., Costa,M.C.R.,
Costa-Neto,C.M., Coutinho,L.L., Cristofani,M., Dias-Neto,E.,
Docena,C., El-Dorri,H., Facincani,A.P., Ferreira,A.J.S.,
Ferreira,V.C.A., Ferro,J.A., Fraga,J.S., Ferreira,A.J.S.,
Frohme,M., Furlan,L.R., Garnier,M., Goldman,G.H., Goldman,M.H.S.,
Gomes,S.L., Gruber,A., Ho,P.L., Hohnsbein,J.D., Junqueira,M.L.,
Kemper,E.L., Kitajima,J.P., Krieger,J.E., Kuramae,E.E., Laigret,F.,
Lambais,M.R., Leite,L.C.C., Lemos,E.G.M., Lemos,M.V.F., Lopes,S.A.,
Lopes,C.R., Machado,J.A., Machado,M.A., Madeira,A.M.B.N.,
Madeira,H.M.F., Marino,C.L., Marques,M.V., Martins,E.A.L.,
Martins,E.M.F., Matsukuma,A.Y., Menck,C.F.M., Miracca,E.C.,
Miyaki,C.Y., Monteiro-Vitorello,C.B., Moon,D.H., Nagai,M.A.,
Nascimento,A.L.T.O., Netto,L.E.S., Nhani Jr.,A., Nobrega,F.G.,
Nunes,L.R., Oliveira,M.A., de Oliveira,M.C., de Oliveira,R.C.,

REMARK
FEATURES
source
gene
CDS
TITLE
JOURNAL
Submitted (02-JUN-2000) Organization for Nucleotide Sequencing and
Analysis, Bioinformatics Lab - IC/Unicamp, C.P. 6176, Campinas, SP
13083-970, Brazil

Location/Qualifiers
1..13856
/organism="Xylella fastidiosa"
/db_xref="taxon:2371"
complement(195..803)
/gene="XF2101"
complement(195..803)
/gene="XF2101"
/note="similar to G11652989 (percent identity: 30 %/query
alignment coverage: 87.6 %/subject alignment coverage:
67.6 %) identified by sequence similarity; putative; ORF
located using Glimmer/RBSfinder."
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAPI1766.1"
/db_xref="GI:9107229"
/translation="MSRCITGRMWNQICVNVVLVFWLVAAGVIWIGQDQOASS
DVIIVLGAADAKPSVFEERIRHGLVLOHGVAPLITFGGFGSSAFARSOVAR
RYALRYDVPRAILIEVYSRTFRLNLOARLMVYGLRAIVVSDPLHARALRLCR
ELIDIALASATPSTRFSLRTMRFLLOEYVFFHRDLVIAN"
complement(800..1042)
/gene="XF2102"
complement(800..1042)
/gene="XF2102"
/note="hypothetical protein; identified by sequence
similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAPI1767.1"
/db_xref="GI:9107230"
/translation="MARVIMYSSFALSVKSCIGVLTGMLLHAYVFGGIGGEPFAL
CCDVAWLCHVDISARLSGCCSAGFICGVSLVVS"
1397..1732
/gene="XF2103"
1397..1732
/gene="XF2103"
/note="hypothetical protein; identified by sequence
similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAPI1768.1"
/db_xref="GI:9107231"
/translation="MKARLFLPILMSAISCATYVNSGYHGPVTEHYHPNYY
PGGCHMDLPNYPYIPYPRDFYRTHLHPYQWRPHHDSRPLPRPQPIKQEQPNIE
MGLRDKQPI"
1914..2024
/gene="XF2104"
1914..2024
/gene="XF2104"
/note="hypothetical protein; identified by sequence
similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAPI1769.1"
/db_xref="GI:9107232"
/translation="NNIADQDMLLCSELELMKICPHCVDTLKHQRLAR"


```

/note="Similar to polynucleotide synthase."
/codon_start=1
/evidence=not_experimental
/protein_id="AAA98699.1"
/db_xref="GI:1293790"
/translation="MDSLSTVDFVNLNKKYEPEDETIASDVEDVPTVYDELADHTVRK
KYSFELHPASGFTMGVNEPIONEEVBOITQSYILKNSPGFKLDYLWISGGLQ
LSNFSSTNLALINDGGQDKELRKLAKRONILINFDLQDSSIEITLYMLINLVKT
LSKSKIKCNFVSGEFLSGNSIARAPMTKISEKYPILSEMDOKIOOVSPDSSSI
PGMWLTIGELSGFEIGKELGNSAEVNLITSRQPSADVTEKRTYATVHTIAD
VNNKKLIEELTRVNDVYGLIHSAGVILKDSKITROTQDSFEVVRPRADQVLEETI
EDGMAITNQYRKFTIADGFLQHNKDALKYILKACIEKPEIVVANAIVAMPYILIKRK
DLPRLDITINGLIFSTVITLKNNGNSQSPFSDNFEVSEINVAEDADVDELLEKYS
ELVCAPIKSKNKNIMDLGMSRLVEFLNLTNLEKIKLNADAVYSLFLEKLA
IPDKLSNTEESPMGODETQETLVDKRETKSTIDCPFGVNAVPDKKNEFKAKSEA
VKCLVSGKTLPPAAGNPAIVIGTSTISDVSKTKATETQVTCQKSKCVMLTGQGS
OYPLMGROLVENEVFRATLQSLCLACKCBETLQGEVSLWELFNTENHYLLQITHMOP
IMPSFGIATQALMSLFGVADYLGHSVGLAAGVLAGIMTLEDGLRILYGRGAMEN
IAGGLALVQREIADYLRKPKVSAIINSKQVYFAGTPELDSALAFVGGKQXG
ITVNOYRPHSSILIDETHLVALRKLADIREFOGATRLVSNVTGQIINTESEAVITKH
TISAVFVDCVETLRNDGVAVMIDAGPAALIAFPKRTIIVPAELSKRITVTCSEKS
DVELTVOCSLELEOGSPFELMPKRYVGCDDTDAPLAHPVQSTIQNHHEELLEGHR
LNGKILIVAGYOMPAMDLYSKASGLDVLKKNKVPWPYIEDCKAFLIRNNDMSI
ELVANSVTYCSAKVEMKBSALETOVIGENKPFVHEVETILYRNGLOYESGEFRMD
SKRSQRCFSSLSKPSFPAMPLIDSAMHSITASVVRPRDSYFLVAMGQVTLKQEDP
KLNNAOAVITVETEDKPIQVNVVAVAGDOICEVRNNTIYVLKIVAEHPPTPTTA
KEAKNSEIIVGFDITLPLYNOLSTNNEMSFLLKTRNXPOLQTNATKEEVALLDD
LRYMDEYFGIRSEAGYLDISOORLLITSAKLKLDLSISTSPKTVFVIGCSTYERS
OIVAHNDENPRAEMGGSTNSALAGRIAHMLKIKGPVLIVDTACSSPHALSACCT
TISGQCECAIVGTVALVHEMTSVLQAOAMTVDOPKAPVDVAGKREAEATLLI
KSTYGLKSMASIVMAVAGHNGNSSLFPNGISOLEIMKATSNTRDIICEVHNGT
KLGDPLELSAISKLSTDGSISSSKNSIGHAEGASGLVICTLIMLSOTYKMAOHL
KCPVITIKGALHCOFGEEDLADBEHSAFLINNGCPGNSCLYTRKQNDPELHESK
DVEYPMLSAHTIKSLERYISVPRFEVASSASIRDMLSLQKVAHVHOPILYSEK
RKLVSODPDDEBNMPTIRORCAVYKEGTVEKATHFQVQLPKPIVNNSLHKLNF
HDEIQOQCVPLREITFEKVLVEKAENNEPMPNVAOVGRLLIPIKTEMEVSCPLNG
ITVPHLSNLSIEELKLELMSLVSPDONSVIIVACFNKGTSHTEMTGVRSLASEP
ILKRVSGIDVHDLKKEPNYQDFESESIFYSNGSRVYERKLRLPTPKMKEITKCLIS
GGTGIGRAVANEIKPKKSVITTRKEKEEPTKSYLSDDVADFSHKPETHVHFKGT
VKNASHLNIKKDILEMYLTFKOGAKNINCCSKLCKFTYSSSIACILGSGFGSNTF
ANGVNSFLNLSRNTKTIHWGPKKEVCLMAKSCDDIKROIEMNGIPLSNDALSV
FSTDLINKHOVYIPDGNFETVSRPHQLMLSEIYLSKTAQKQSPDESFEYVU
GITRMDNKKHILPEMDLIGLSCMENLRHTLNLNLEINLVSEIENATYKLEKLYAS
MLKSTSEKSPKSDNITNRENDENRVAIVAGSAEENIEEEMENLIDGVYATG
KENSLSKPEGFENKFNITDEDAKMLDPOVRKTIOMSTYALNENSGYVRODLKAY
FAGSESDYGRPESONDAMKRLPYMNNNTLASASTACDLKGETVSVASCTITVA
VANAOKTIOGSDVYALVGTASIAEASECISGNSDEKTIFFSKSGICRPRDSSEIV
RGSVGCIVLKYIKSEAKKNDNDNHFVILKDAISNDGISRASFPAMPADQOQKMEVL
EKLQEBENRLEVEVCHATGTTLGDITENNALQTAOCFKKNIQISGKANTIGHAVAS
GLASLICKAKMLQTKVAPOVNFSEPRDSFGTFPNSTOKSFIONDSFIAIDSGIGG
TNVHMIIERLIGIREKNDDEVENVILPISANSKASLVEAIKTMEVOLKKSNIKLSST
LLHHRNOMNEFAITISKIIGKVOIKSFKIKTMSKSPKALFPFAGIOGPIANILPHE
YVHNFDEKSLERHCKTAHILGVSNLLEITPMTYENEDLIDITLTSQSLSCVQMAI
FECLAVIKPOLIIGHSVGEYAAVAVISQCSIESEMELLERSNLIISQYOKAMAI
NYREELDNSVYSIAIDITIKCIVGPEALIDELEKTFKCOVKYKNTAKNINHHFHKML
QVSOKEKNLCLSPSEPRPIPIISSITGQRIKRPFAEYCRKHLITVAIVLDVNNIT
KSGDVAYVEIGPTGVSLNLGKKNKSIIVPTGCTGKHPKASISDCTIGLWCKGDIR
PLPEKLIKIDENAGVCEDEYVPGHKKTPVYKFNYSKSKKNNKSNSTETALASDFQ
LYDNKISOTQSTILYVHLHGCSDSHONNFEVVKPIKQCYETICEITTSDSASVHI
IGLIRHQIYSPRIKIKVFNEMLELKAVNSAIASDGLYVAVTSEGVHEFLDISP
NTHSIKTKMEKMLVFGASGVGOSARLHSGNVAIEVSSRSKISDRIHLKASQVE
VLSNLKHFVYLVNCVGEKSSSKINKTFEBOEETLITVSGNSNILKALQOESK
KYITFSSLSVILGNSDYASANCVEALKSTFVOKLVLVSLPLVJDSRWKERTES
TKDMLSNLSNNEELNALKIALTEEGTFLVLEDDPAOIAKSLSHRDEIGEV
ENOBRECLVTSIRINIKETLIGOSIGTIDISNPSIGGDSLSALQVWQVOKTGTTE
TSPYNIFFELVAVYDIFSCWNSLKYTSLIAYQTSYRSVFNPSRPLOKOKSDESM
HNLDMDRLDAISQELYSFETGQTPPLRVIAEDSDNNOIHVFNQHILITDEKMS
TVLADTISKLYTSFKLTVPLKSNQSVAKLAIQOCVENLQPEAVNAKSEYENHPTI
IPYDVAVEENHKEFKIRIPIRINISKIYADMSKYSTTTHNCLAVPECCYASFTGQI
DLMAFANSGRNSENKDIYGVPMNNALEFSTISEVBEFEKLTIVSVSLDCGRCSN
VPHPTIIVEONRKINAIISTYFNFRQKLDIPSVSIPGATCEIKHLISLNNAFSPSTIDET

```

gene

CDS

tRNA

gene

CDS

gene

CDS

gene

CDS

```

PTGPILSIYDSSKYPFESTVNLPSFELPKAKKFSKQKNNKIIHRRRADPSPFLQ
IDLITSMKLPDSDESPALLSASTIMYSELALIEKVAACIKQVQIEEASVAREDDIT
GVYSSKTPISILACSLIGASVAPVPSDEQOLYKSVINIVEDSDHVEFLNLR
SEKTRQYAGIYITFSGSTGADKGVFAAGQSVSEFLSATROCKCFRERKVDLSVH
VFDVSYNIFGPMILNGCCISADHSTTITDLOKQOYAFILPAAPNFGTSETTARLRT
VENTLITGGEFVSADALOCALKMYPRMIRIIOITGPIETCVMSLJINCRITSEENGANI
IGITIDQVAKARAVDLELEKELLCGLQVLAIVYNNQMLIAFLAGTSTAFQYSDI
IALQKRTLPHHYVOIDQMPLNSGKIEKSTLLQAFESVYSKREVPKNSLEEK
IYASGDVIGRTELEIDKESDVGNLSLAIOLAHLELKEVDYKAHEHIIOSNNGCL
SKLASVSPKPSKPNVTTKLREVPNSKFSIYVLAHIGTITTYPSFLOTPRNNISL
YGLEPDKYPCGDLREILAHFYQELIAAAGOKOIFVGHSGGIMSREIVAEKILWY
OIFEWLPDSWVLRTEINDIENTIKOFTIVYFSSGLPDSHRIDRAIKLQALREKTSY
SDTKLIRSKQDGLDAARKKAVRDLNELSRMTQNGFDELSTQPIETYLIDGDHES
CLKVENLKVKDPIILAPKPYDDA"
1813. .19760
/gene="G01D9.4"
/join(1813. .18387,18490. .18569,18615. .18737,18791. .18886,
19060. .19237,19285. .19395,19444. .19624,19684. .19760)
/gene="G01D9.4"
/codon_start=1
/evidence=not_experimental
/protein_id="AAA98700.1"
/db_xref="GI:1293791"
/translation="MTITKTEPTDCKNAEGYEDNPVKYVCEADEEKHISNGVWP
MEGWFDPDKERSSSEGLSIIIRLSPPILARPPSPDPYEANAVARRRKCTKLE
TLIAFESVILVIFELVYQGLISVAYQKGVSGMAMFNRQTELELVONVIMND
SEKIDVPEFGSNRPALFLHDFKONLAIYDVGNCRFVYDLDRTKIRSPRMILELR
NIDIMAPVAAVQOTRMVRETVNGVDELTKADISTFNSTILSRICARQYTKIKLKQDAQ
TRSRRAACMGALISFASMAQSYVLEELSF"
complement(19138. .19236)
/note="Leu: codon recognized: CTA"
20621. .21971
/gene="G01D9.3"
/join(20621. .20772,21464. .21614,21664. .21783,21834. .21971)
/gene="G01D9.3"
/codon_start=1
/evidence=not_experimental
/protein_id="AAA98701.1"
/db_xref="GI:1293792"
/translation="MKIFNLSPFIVISTTRKPPCCGRHLGAAACSKIMHONTLPKARC
NSDAFPLIOCCSCNTDGYGMAVDIARLSVSEKFPDYGFCORLSPPLSNTP
HNTSCGGENPQIATFRCKRSCGYCNFSVQYTLDNAQACRVQPAEKKRMWKYFV
SSTITTSBSPMEVINSTYQFNHN"
complement(22269. .22799)
/gene="G01D9.6"
complement(join(22269. .22371,22415. .22541,22588. .22680,
22730. .22799))
/gene="G01D9.6"
/codon_start=1
/evidence=not_experimental
/protein_id="AAA98702.1"
/db_xref="GI:1293793"
/translation="MIPRFLIKYICLISLIGPVYDASTSMYRRRTPSERWASIR
PRMFOPEKTIQOTITOLDNRRALPCKCKDESGSICKNLKRTDILFLYQKQTEP
DESLVOCNCSAGISTYKRRKQOVNL"
32520. .36916
/gene="G01D9.2"
/join(32520. .32678,33205. .33505,33554. .33662,33836. .33936,
33983. .34088,34698. .34775,34824. .35280,35332. .35502,
35548. .35847,36442. .36637,36699. .36916)
/gene="G01D9.2"
/note="Similar to receptor tyrosine kinase."
/codon_start=1
/evidence=not_experimental
/protein_id="AAA98703.1"
/db_xref="GI:1293794"
/translation="MONGVDLDAQITASSSDKQSVGQONARLHSELASGAMCPKPOL
NKSSTFEFLVNDITLFTSVTQRYNGKREASHYMDIARLPSQOMVIRYLNRSR
HYMDGNPOTTPVIRALDPPIVASRIINVSCKRTVCARAEIKHKGEGVLYYSG
VDSGRILDLDFKDMNFENSOYTESGTLRGLLIDVYVQTSPPFKNOMNNSNWCIO
NBDTGDGAVTILFEFEIHNFDVLAFFGNRIGLDIVYSQDKTPTPLSQISSSR
QTLNNTSRNYDRVPLNHRACKKVAISIKFSDDMKFLTEVHFTSGITRHTVSOKNINY

```



```

gene
/translation="MWTCLAGAGVLEGRATHVLLTSLRGROFTLLPALLFPCPL
GCMGLIAVLYKDNIEFQEAVERKCEVPCVGNKMAKSTVEVGAERWRLLEYI
GVLILPILSLVFCQKLLMFLGACVSVREKAKGLDRAPCFPLSLIMNPDLGKC
NMHPELKIS"
join(37532, .38415,38988, .39218,39280, .39471,39552, .39890,
40001, .40207,40381, .40515,40601, .40765)
/gene="P0686E09.6"
join(37532, .38415,39005, .39206,38988, .39218,39280, .39471,
39552, .39890,40001, .40207,40381, .40515,40601, .40765)
/feature="hypothetical protein
similar to Oryza sativa chromosome 1, P0665D10.16"
/codon_start=1
/protein_id="BAB39114.1"
/db_xref="GI:13365569"
/translation="MSEADRLRRAAALALADGGGAGRDPAKADAVDFADPSPV
LRASVATSSSSSSGSAKSPAPNAGALADGRSHSGELTAPTRPGRRCSS
GPLIFSGSGSGGGGDRGTASPTNLAIPGNCISGRVAVAAAPPPSRD
VLGSGTGYGHSIMRGSGMAPARSSIDSSFLGAPRSPATPRASSASGSLDQY
TRICNMYKKKKKHAALRHYDRVALCPESACGNNAALAGRLADLRCEERY
RLDPANGRAHRLAGLRLGMLISKARHLTQAGHLHSDSEMERLQEVHOGRSI
DARKVGWKSALREDAALAGADSSRLTLGLMISARHLTQAGHLHSDPSE
MEKLOEVMHOGRSIDARKVGDMSALREDAALAGADSSRLTMSLIRSFALR
LHLEFADSTLAKLDSVLYIRMGANPSGMLAESVYIRAVDMALRCRFDRA
VEADNARFIDPNAEYGMILNNYKLVAKAOGNELYKAKFSADATVASEGICYEP
SNPLYCNBAACWGLERMEKAVDDCNALRIOPNTYKALIRASVYAKVERADCYR
DYEVLEKIPADPREVAELRFAOYALKTGTGEDVSNKREGVEMYTSVQLRAIIS
PGVLEORAISSGVVYFMSTMOQCTLLTPSVNSLSECPLELTVQVNDSPV
AKAENRIVLPFKIKYKGVKVKEMICPSLHVLRVHYVSS"
complement(join(47016, .47264,47291, .47363,48616, .48748,
48784, .48832,49178, .49324,49569, .49684,50921, .50967,
51070, .51107,51434, .51652,51763, .51861,51888, .51986,
52059, .52112,52817, .52933,53291, .53532,54642, .54895,
55164, .55345))
/gene="P0686E09.7"
complement(join(47016, .47264,47291, .47363,48616, .48748,
48784, .48832,49178, .49324,49569, .49684,50921, .50967,
51070, .51107,51434, .51652,51763, .51861,51888, .51986,
52059, .52112,52817, .52933,53291, .53532,54642, .54895,
55164, .55345))
/feature="hypothetical protein
similar to Arabidopsis thaliana chromosome 3, MBK21.5"
/codon_start=1
/protein_id="BAB39115.1"
/db_xref="GI:13365570"
/translation="MSDRVSQRENTMDMGYARENKAMDGLIAYVEGARSCKCA
LGGARVHILIGRVOGSASQPPSAKTLASASGASMSDPTPHCGGAGBEGA
AAGTSSTEVILIAKRTGRPLANPSAADLADAGARSMCVLYHQPMGSHSTVE
QIEALYSNTMTKARPLESSLPIYPPORGEKDAFYMMTRKCEVLCRKHDSQWPE
GGIPNMKEHLFEALDNTLLIKGYIGNSCOHAFVYLTLAGQPRAAVDESYDEO
EGEDPCPFPMKGTGCKGSKCKENHPEKYNALASGNTNDTNNSSOPLRSQPCSF
MKLOKHLIADSSILPEYAKGCKGCKENHPEKYNALASGNTNDTNNSSOPLRSQPCSF
GSAADSVASAKMOPVAAOEFNSKGLPMRGEVDCPRYMGSCCKGCTGCRHREHRL
VLRNPLPGOTIILPTPSMLINSANPMQGFDRHAAMPYMTGTCKRFDKCRFHPR
IDRSADPPSAMPEAESVOLITLAGLPRREDVAVCAVMTGCKGEGYGVSEFEL
RRETNILJGCRISMPYSFLAYELNMKPHYLGSFLYKRRPLSELESPPKAAALISA
QPVAYAGHRLRLAQAVALPHPTASGREPIITRGSALATMSPRRAAGGDAVRLGCLPV
GCAAMPARSAASGDHPLCCAGY"
join(57612, .57759,58207, .58289,58617, .58775,59266, .59373,
59488, .59548,59653, .59918,60187, .60219)
/gene="P0686E09.8"
join(57612, .57759,58207, .58289,58617, .58775,59266, .59373,
59488, .59548,59653, .59918,60187, .60219)
/feature="contains ESTs D40230(S2054),D23027(C2064)
hypothetical protein
similar to Arabidopsis thaliana chromosome 1, F19K6.5"
/codon_start=1
/protein_id="BAB39116.1"
/db_xref="GI:13365571"
/translation="MVGMAAGARAVLRRLSAAEAQAAGKODGRVFAASYSGGGPN
APRIGOVANILRAOAFASRGVALNAGLSTTRNLAAADAMVPVSSPLTPR
LGSGEOTDKKGAIVKRLKVOAIKKDKIKQSPKKVNLAKLRCGRVEDALLQLOVYTKR

```

AACTVYSARANAHHGLDPPKLIYEAFVSGKLYLRLSYHANGRCGVWPRCRL
TVVVRATAEERAKIAKILRVSNYKLTREKOLMPHRLIEVSPRARRRKEEGAAAL

Query Match 64.0%; Score 16; DB 13; Length 156425;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 52932 ATCGCTGTTGAAGC 52917

RESULT 28

AC016697/c

LOCUS

DEFINITION

AC016697.7

AC016697.7

AC016697.7

AC016697.7

AC016697.7

AC016697.7

AC016697.7

AC016697.7

AC016697.7

AC016697.7

AC016697.7

AC016697.7

AC016697.7

AC016697.7

AC016697.7

AC016697.7

AC016697.7

AC016697.7

AC016697.7

AC016697.7

AC016697.7

AC016697.7

AC016697.7

AC016697.7

AC016697.7

AC016697.7

AC016697.7

AC016697.7

AC016697.7

AC016697.7

AC016697.7

AC016697.7

AC016697.7

AC016697.7

AC016697.7

AC016697.7

AC016697.7

AC016697.7

AC016697.7

AC016697.7

AC016697.7

AC016697.7

AC016697.7

AC016697.7

AC016697.7

AC016697.7

AC016697.7

AC016697.7

AC016697.7

AC016697.7

AC016697.7

AC016697.7

AC016697.7

AC016697.7

AC016697.7

AC016697.7

AC016697.7

AC016697.7

```

misc_feature      1..642
                  /note="assembly_name:Contig20"
misc_feature      743..164227
                  /note="assembly_name:Contig21"
BASE COUNT      46978 a 33168 c 34928 g 49053 t      100 others
ORIGIN
Query Match      64.0%; Score 16; DB 64; Length 164227;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      7 gcctgttgagcgagc 22
        |||
Db      40943 GCTGTTTGAAGCGCAG 40928

RESULT 29
AC023914/c      AC023914 168177 bp DNA HTG 20-SEP-2000
LOCUS
DEFINITION      Homo sapiens clone RP11-44K20, WORKING DRAFT SEQUENCE, 22 unordered
                  pieces.
ACCESSION      AC023914
VERSION      AC023914.3 GI:10198362
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS      Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
              Anderson, S., Baldwin, J., Barna, N., Beda, F., Boguslavsky, L.,
              Boucknight, B., Brown, A., Burkett, G., Campopiano, A., Castle, A.,
              Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,
              Dastellano, K., Dewar, K., Dodge, S., Domino, M., Doyle, M.,
              Fenesstor, J., Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D.,
              Galaan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
              Grand-Pierre, N., Grant, G., Hages, B., Heaford, A., Horton, L.,
              Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
              Klein, J., Landers, T., Largocque, K., Lehoczky, J., Levine, R.,
              Liew, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McCarthy, M.,
              McEwan, P., McGuck, A., McKernan, K., McPheters, R., Meldrum, J.,
              Menaus, L., Mihova, T., Miranda, C., Mlenka, V., Morrow, J., Naylor, J.,
              Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Olivar, T.M.,
              Peterson, K., Pierre, N., Pisani, C., Pollara, V., Raymond, C.,
              Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S.,
              Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
              Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Thirell, A.,
              Travers, M., Trifilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B.,
              Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zimmer, A. and
              Zody, M.
DIRECT SUBMISSION
Submitted (20-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Sep 20, 2000 this sequence version replaced gi:9102875.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www.seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu

Project Information
Center project name: L1087
Center clone name: 44_K_20
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads

```

```

FEATURES
source
1..168177
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="RP11-44K20"
/clone_lib="RP11-44K20"
/misc_feature
1..1263
/note="assembly_fragment"
clone_end:sp6
vector_side:left"
misc_feature
1364..2399
/note="assembly_fragment"
2500..3713
misc_feature
1263: contig of 1263 bp in length
1264 1363: gap of 100 bp
1364 2399: contig of 1036 bp in length
2400 2499: gap of 100 bp
2500 3713: contig of 1214 bp in length
3714 3813: gap of 100 bp
3814 27125: contig of 23312 bp in length
27126 27225: gap of 100 bp
27226 28854: contig of 1629 bp in length
28855 28954: gap of 100 bp
28955 31625: contig of 2671 bp in length
31626 31725: gap of 100 bp
31726 34084: contig of 2359 bp in length
34085 34184: gap of 100 bp
34185 36582: contig of 2398 bp in length
36583 36682: gap of 100 bp
36683 39629: contig of 2947 bp in length
39630 39729: gap of 100 bp
39730 43900: contig of 4171 bp in length
43901 44000: gap of 100 bp
44001 46837: contig of 2837 bp in length
46838 46937: gap of 100 bp
46938 52599: contig of 5662 bp in length
52600 52699: gap of 100 bp
52700 60235: contig of 7536 bp in length
60236 60335: gap of 100 bp
60336 67600: contig of 7265 bp in length
67601 67700: gap of 100 bp
67701 76219: contig of 8519 bp in length
76220 76319: gap of 100 bp
76320 85003: contig of 8684 bp in length
85004 85103: gap of 100 bp
85104 95875: contig of 10772 bp in length
95876 95975: gap of 100 bp
95976 108080: contig of 12105 bp in length
108081 108180: gap of 100 bp
108181 126996: contig of 18816 bp in length
126997 127096: gap of 100 bp
127097 145024: contig of 17928 bp in length
145025 145124: gap of 100 bp
145125 166554: contig of 21330 bp in length
166555 166554: gap of 100 bp
166555 168177: contig of 1523 bp in length.
Location/Qualifiers
1..168177
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="RP11-44K20"
/clone_lib="RP11-44K20"
/misc_feature
1..1263
/note="assembly_fragment"
clone_end:sp6
vector_side:left"
misc_feature
1364..2399
/note="assembly_fragment"
2500..3713

```

```

NOTE: This is a 'working draft' sequence. It currently
consists of 22 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

```



```

* 177771 178025: contig of 255 bp in length
* 178026 178045: gap of unknown length
* 178046 178482: contig of 437 bp in length
* 178483 178502: gap of unknown length
* 178503 180037: contig of 1535 bp in length
* 180038 180057: gap of unknown length
* 180058 180950: contig of 893 bp in length
* 180951 180970: gap of unknown length
* 180971 182851: contig of 1881 bp in length
* 182852 182871: gap of unknown length
* 182872 183253: contig of 382 bp in length
* 183254 183273: gap of unknown length
* 183274 183953: contig of 680 bp in length
* 183954 183973: gap of unknown length
* 183974 184181: contig of 208 bp in length
* 184182 184201: gap of unknown length
* 184202 185477: contig of 1276 bp in length
* 185478 185497: gap of unknown length
* 185498 186883: contig of 1386 bp in length
* 186884 186903: gap of unknown length
* 186904 187053: contig of 150 bp in length
* 187054 187074: gap of unknown length
* 187074 188199: contig of 1126 bp in length.
location/Qualifiers
1. 188199
/organism="Mus musculus"
/strain="C57BL6/J"
/db_xref="taxon:10090"
/chromosome="18"
/clone="RP23-233F17"
/sex="male"
1. 34856
/note="assembly_name:Contig108"
34877. 63119
/note="assembly_name:Contig107"
63140. 78565
/note="assembly_name:Contig106"
78586. 94535
/note="assembly_name:Contig105"
clone_end:SP6
vector_side:right"
94556. 112689
/note="assembly_name:Contig104"
112710. 124902
/note="assembly_name:Contig103"
clone_end:T7
vector_side:left"
124923. 125260
/note="assembly_name:Contig102"
125281. 137456
/note="assembly_name:Contig101"
137477. 145685
/note="assembly_name:Contig100"
145706. 152451
/note="assembly_name:Contig99"
152472. 159495
/note="assembly_name:Contig98"
159516. 162862
/note="assembly_name:Contig97"
162883. 166492
/note="assembly_name:Contig96"
166513. 169948
/note="assembly_name:Contig95"
169969. 173608
/note="assembly_name:Contig94"
173629. 175468
/note="assembly_name:Contig93"
175489. 176715
/note="assembly_name:Contig92"
176736. 177521
/note="assembly_name:Contig91"
177542. 177750
/note="assembly_name:Contig90"

```

```

misc_feature 177771. 178025
/note="assembly_name:Contig89"
misc_feature 178046. 178482
/note="assembly_name:Contig88"
misc_feature 178503. 180037
/note="assembly_name:Contig87"
misc_feature 180058. 180950
/note="assembly_name:Contig86"
misc_feature 180971. 182851
/note="assembly_name:Contig85"
misc_feature 182872. 183253
/note="assembly_name:Contig84"
misc_feature 183274. 183953
/note="assembly_name:Contig83"
misc_feature 183974. 184181
/note="assembly_name:Contig82"
misc_feature 184202. 185477
/note="assembly_name:Contig81"
misc_feature 185498. 186883
/note="assembly_name:Contig80"
misc_feature 186904. 187053
/note="assembly_name:Contig79"
misc_feature 187074. 188199
/note="assembly_name:Contig78"
BASE COUNT 51323 a 42282 c 41293 g 52561 t 740 others
ORIGIN

```

```

Query Match 64.0%; Score 16; DB 78; Length 188199;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 ggaatcgcgtttgaa 16
|||||
Db 108548 GGATCTGCTTTGAA 108563

```

```

RESULT 31
LOCUS AC007104/c DNA 17-JUL-1999
DEFINITION Homo sapiens Chromosome 4, *** SEQUENCING IN PROGRESS ***, 16
unordered pieces.
ACCESSION AC007104
VERSION AC007104.4 GI:5523795
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 200000)
Stone,N.E., Schmutz,J.J., Cox,D.R. and Myers,R.M.
Direct Submission
Unpublished
2 (bases 1 to 200000)
Stone,N.E., Schmutz,J.J., Cox,D.R. and Myers,R.M.
Direct Submission
REFERENCE 1 (bases 1 to 200000)
AUTHORS Stone,N.E., Schmutz,J.J., Cox,D.R. and Myers,R.M.
TITLE Submitted (17-MAR-1999) Department of Genetics, Stanford Human
JOURNAL Genome Center 855 California Avenue, Palo Alto, CA 94304, USA
on Jul 17, 1999 this sequence version replaced gi:4757683.
COMMENT
* NOTE: This is a 'working draft' sequence. It currently
* consists of 16 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1369: contig of 1369 bp in length
* 1370 3774: gap of unknown length
* 3775 4878: contig of 1104 bp in length
* 4879 7283: gap of unknown length
* 7284 8555: contig of 1272 bp in length
* 8556 10960: gap of unknown length

```

```

* 10961 12207: contig of 1247 bp in length
* 12208 14612: gap of unknown length
* 14613 16262: contig of 1650 bp in length
* 16263 18667: gap of unknown length
* 18668 20176: contig of 1509 bp in length
* 20177 22581: gap of unknown length
* 22582 24278: contig of 1697 bp in length
* 24279 26684: gap of unknown length
* 26684 28199: contig of 1515 bp in length
* 28199 30604: gap of unknown length
* 30604 32359: contig of 1756 bp in length
* 32359 34764: gap of unknown length
* 34764 39958: contig of 5194 bp in length
* 39959 42363: gap of unknown length
* 42363 57085: contig of 14722 bp in length
* 57086 59490: gap of unknown length
* 59491 74905: contig of 15415 bp in length
* 74906 77311: gap of unknown length
* 77311 84273: contig of 6963 bp in length
* 84274 86678: gap of unknown length
* 86679 118061: contig of 31383 bp in length
* 118062 120466: gap of unknown length
* 120467 152965: contig of 32499 bp in length
* 152966 155371: gap of unknown length
* 155371 200000: contig of 44630 bp in length.

```

```

FEATURES
  source          1..200000
                  location/Qualifiers
                  /organism="Homo sapiens"
                  /db_xref="taxon:9606"
                  /chromosome="4"

```

```

BASE COUNT      38319 a 42757 c 43367 g 37466 t 38091 others
ORIGIN

```

```

Query Match
Best Local Similarity 64.0%; Score 16; DB 60; Length 200000;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      3 atctgctgttgaag 18
        |||
Db 129264 ATCTGCTGTTGAAGC 129249

```

```

RESULT 32
LOCUS   PEHE7D      306 bp      DNA      VRL      10-FEB-1999
DEFINITION Human papillomavirus type 21 DNA for E7 protein, complete cds.
ACCESSION D50548.1 GI:808883
VERSION   D50548.1 GI:808883
KEYWORDS  E7 protein.
SOURCE    Human papillomavirus type 21 DNA.
           Human papillomavirus type 21
           Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
           Papillomavirus.
REFERENCE 1 (bases 1 to 306)
AUTHORS   Kiyono,T.
TITLE     Direct Submission
JOURNAL   Submitted (11-MAY-1995) to the DDBJ/EMBL/Genbank databases. Tohru
           Kiyono, Aichi Cancer Center, Research Institute, Laboratory of
           Viral Oncology; 1-1 Kanokoden, Chikusa-ku, Nagoya, Aichi 464, Japan
           (E-mail: h44714@nuc.cc.nagoya-u.ac.jp, Tel: 052-762-6111(ex.8838),
           Fax: 052-763-5233)
REFERENCE 2 (bases 1 to 306)
AUTHORS   Adachi,A., Kiyono,T., Ohashi,M. and Ishibashi,M.
TITLE     Detection of human papillomavirus type 47 DNA in malignant lesions
           of epidermodysplasia verruciformis with protocols for precise
           typing of the related HPV DNAs
JOURNAL   unpublished (1995)
REFERENCE 3 (sites)
AUTHORS   Aachhi,A., Kiyono,T., Hayashi,Y., Ohashi,M. and Ishibashi,M.
TITLE     Detection of human papillomavirus (HPV) type 47 DNA in malignant
           lesions from epidermodysplasia verruciformis by protocols for
           precise typing of related HPV DNAs

```

```

JOURNAL      J. Clin. Microbiol. 34 (2), 369-375 (1996)
MEDLINE      96381005
FEATURES
  source      Location/Qualifiers
              1..306
              /organism="Human papillomavirus type 21"
              /db_xref="taxon:31548"

```

```

gene         1..306
             /gene="E7"
CDS          1..306
             /gene="E7"
             /note="putative"
             /codon_start=1
             /product="E7 protein"
             /protein_id="BA09116.1"
             /db_xref="GI:808884"
             /translation="MIGKEVTLQDILVELNELQPEVOPVDLFCHEELPSQOEETFEEL
             PRTSKVVTGCGCKVKRLRFVNTFRFVNLRFQNLRFELQLLCPEDCRGNCKIGGS"

```

```

BASE COUNT      88 a 59 c 81 g 78 t
ORIGIN

```

```

Query Match
Best Local Similarity 60.0%; Score 15; DB 59; Length 306;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      3 atctgctgttgaag 17
        |||
Db      236 ATCTGCTGTTGAAG 250

```

```

RESULT 33
LOCUS   PEU244276      625 bp      DNA      PLN      11-MAY-1999
DEFINITION Pringsheimia euphorbiae 5.8S rRNA gene and internal transcribed
ACCESSION AJ244276
VERSION   AJ244276.1 GI:4808321
KEYWORDS  5.8S ribosomal RNA; 5.8S rRNA gene; internal transcribed spacer 1;
           internal transcribed spacer 2; ITS1; ITS2.
SOURCE    Pringsheimia euphorbiae.
           Pringsheimia euphorbiae.
           Eukaryota; Fungi; Ascomycota; Pezizomycotina; Dothideomycetes et
           Chaetothyriomycetes incertae sedis; Dothioraceae; Pringsheimia.
REFERENCE 1 (bases 1 to 625)
AUTHORS   Yurlova,N., Hoog,G.D. and Gerrits van den Ende,A.
TITLE     Taxonomy of Aureobasidium and allied genera
JOURNAL   Unpublished
REFERENCE 2 (bases 1 to 625)
AUTHORS   Hoog,G.D.
TITLE     Direct Submission
JOURNAL   Submitted (03-MAY-1999) Hoog G.D., Research, Centraalbureau voor
           Schimmelcultures, P.O.Box 273, 3740 AG Baarn, NETHERLANDS

```

FEATURES

```

  source      1..625
              Location/Qualifiers
              /organism="Pringsheimia euphorbiae"
              /strain="CBS 747.71"
              /db_xref="taxon:93493"
              <1..196
              /note="internal transcribed spacer 1, ITS1"
  misc_feature 197..351
              /gene="5.8S rRNA"
              /product="5.8S ribosomal RNA"
  rRNA         197..351
              /gene="5.8S rRNA"
  gene         352..>625
              /note="internal transcribed spacer 2, ITS2"
  misc_feature 170 a 128 c 135 g 192 t
BASE COUNT      170 a 128 c 135 g 192 t
ORIGIN

```

```

Query Match
Best Local Similarity 60.0%; Score 15; DB 14; Length 625;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```


gene
/organism="Serratia marcescens"
/db_xref="taxon:615"
656..1465
/gene="aac3-Vb"
656..1465
/gene="aac3-Vb"
/codon_start=1
/transl_table=11
/product="acetyltransferase"
/protein_id="AA26548.1"
/db_xref="GI:152812"

BASE COUNT 276 a 516 c 506 g 274 t
ORIGIN

Query Match 60.0%; Score 15; DB 3; Length 1572;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 ctgttgaagcgag 22
|||||
Db 1397 CTGTTGAGCGCAG 1411

RESULT 37
AF239157/c 1616 bp mRNA ROD 13-MAR-2000
LOCUS Rattus norvegicus DEXRAS1 (Dextrasi) mRNA, complete cds.
DEFINITION AF239157
ACCESSION AF239157.1 GI:7230767
KEYWORDS
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 1616)
Fang, M., Jaffrey, S.R., Sawa, A., Ye, K. and Snyder, S.H.
Dextrasi: A G-protein coupled to neuronal nitric oxide synthase via
CAPON
Unpublished
2 (bases 1 to 1616)
Fang, M., Jaffrey, S.R., Sawa, A., Ye, K. and Snyder, S.H.
Direct Submission
Submitted (24-FEB-2000) Neuroscience, Johns Hopkins University, 725
N. Wolfe St, Baltimore, MD 21205, USA
Location/Qualifiers
1..1616
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
1..1616
/gene="Dextrasi"
118..960
/gene="Dextrasi"
/note="Ras-related protein"
/codon_start=1
/product="Dextrasi"
/protein_id="AA26548.1"
/db_xref="GI:7230768"
/translation="MKLAAMTKMKCPDSDELSPANKCYRMVILGSSKVGKTAIVSRF
LTGRFDAYPTIEDFHRKFSYIRGEVYQDIDTSGNHPPAMRLSTITGDVFLV
FSLDNRDSFEVORLKOQIIDTKSCIKNKTKENVDPVLYICGNKGRDPTREYREOREI
EQVLGDDPCRCAYFEISAKKNSSLDOMFRALFPMALPSEMSPDHRRKVSVOYCVLH
KKALRNKKLIRAGSGGGGHDGDAFGILAPPARPSVSHDLMIYRRTSVSSQAKKER
CVIS"

BASE COUNT 403 a 436 c 413 g 363 t 1 others
ORIGIN

Query Match 60.0%; Score 15; DB 94; Length 1616;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ggaatcgtgttga 15
|||||
Db 481 GGATCTGCTGTTGA 467

RESULT 38
AF009246/c 1623 bp mRNA ROD 13-FEB-1998
LOCUS Mus musculus ras-related protein (DEXRAS1) mRNA, complete cds.
DEFINITION AF009246
ACCESSION AF009246.1 GI:2253712
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1623)
Kempainen, R.J. and Behrend, E.N.
Dexamethasone rapidly induces a novel ras superfamily
member-related gene in AtT-20 cells
J. Biol. Chem. 273 (6), 3129-3131 (1998)
2 (bases 1 to 1623)
Kempainen, R.J. and Behrend, E.N.
Submitted (18-JUN-1997) Physiology & Pharmacology, Auburn
University College of Veterinary Medicine, 213 Greene Hall, Auburn,
AL 36849, USA
Location/Qualifiers
1..1623
/organism="Mus musculus"
/db_xref="taxon:10090"
/cell_line="corticotrope tumor cell line AtT-20"
1..1623
/gene="Dextrasi"
142..984
/gene="Dextrasi"
/note="induced by dexamethasone"
/codon_start=1
/product="ras-related protein"
/protein_id="AAC5358.1"
/db_xref="GI:2253713"
/translation="MKLAAMTKMKCPDSDELSPANKCYRMVILGSSKVGKTAIVSRF
LTGRFDAYPTIEDFHRKFSYIRGEVYQDIDTSGNHPPAMRLSTITGDVFLV
FSLDNRDSFEVORLKOQIIDTKSCIKNKTKENVDPVLYICGNKGRDPTREYREOREI
EQVLGDDPCRCAYFEISAKKNSSLDOMFRALFPMALPSEMSPDHRRKVSVOYCVLH
KKALRNKKLIRAGSGGGGHDGDAFGILAPPARPSVSHDLMIYRRTSVSSQAKKER
CVIS"

BASE COUNT 412 a 437 c 417 g 357 t
ORIGIN

Query Match 60.0%; Score 15; DB 94; Length 1623;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ggaatcgtgttga 15
|||||
Db 505 GGATCTGCTGTTGA 491

RESULT 39
AF039572 2795 bp DNA BCT 02-DEC-1998
LOCUS Serratia liquefaciens serrawetlin synthase (swrA) gene, partial
DEFINITION AF039572
ACCESSION AF039572
VERSION AF039572.1 GI:2760927

SOURCE	Human papillomavirus type 21.
ORGANISM	Human papillomavirus type 21
REFERENCE	Viruses; dsDNA viruses, no RNA stage: Papillomaviridae;
AUTHORS	Papillomavirus.
TITLE	1 (bases 1 to 7779)
JOURNAL	Delius, H.
REFERENCE	Sequenced by Hajo Delius, Deutsches Krebsforschungszentrum, Angewandte Tumorstrologie, I.N.F. 506, W-6900 Heidelberg, Germany
AUTHORS	Unpublished
TITLE	2 (sites)
JOURNAL	Kremsdorf, D., Favre, M., Jablonska, S., Obaiek, S., Rueda, L.A.,
REFERENCE	Lutzner, M.A., Blanchet-Bardon, C., Van Voortst Vader, P.C., and Orth, G.
AUTHORS	Molecular cloning and characterization of the genomes of nine newly
TITLE	recognized human papillomavirus types associated with
JOURNAL	epidermodysplasia verruciformis
REFERENCE	J. Virol. 52 (3), 1013-1018 (1984)
AUTHORS	85033930
TITLE	3 (sites)
JOURNAL	Kiyono, T., Hiraiwa, A. and Ishibashi, M.
REFERENCE	Differences in transforming activity and coded amino acid sequence
AUTHORS	among E6 genes of several papillomaviruses associated with
TITLE	epidermodysplasia verruciformis
JOURNAL	Virology 186 (2), 628-639 (1992)
REFERENCE	92124737
AUTHORS	4 (bases 1 to 7779)
TITLE	Farmer, A.D.
JOURNAL	Direct Submission
REFERENCE	Submitted (18-JUL-1995) Andrew D. Farmer, HIV Sequence Database,
AUTHORS	Los Alamos National Laboratory, T-10, Mail Stop K710, Los Alamos,
TITLE	NM 87501, USA
COMMENT	HPV21 was originally isolated from skin warts of an
FEATURES	epidermodysplasia verruciformis (EV) patient [2]. Hybridization
SOURCE	assays and phylogenetic reconstructions based on DNA sequences
location/Qualifiers	indicate that HPV21 is most closely related to HPV14 and HPV20, and
1. .7779	then to HPV19 and HPV25. This grouping agrees with assays of the
/organism="Human papillomavirus type 21"	degree of transforming activity of the E6 gene (these related HPV
/db_xref="taxon:31548"	types had relatively low transforming activity as compared to HPV
/note="cloned HPV-21 was obtained from the Papillomavirus	5, 8, and 47), and clustering of similarity of amino acids in the
Reference Center, Heidelberg and sequenced by Dr. H.	second zinc finger domain of E6 [3]. The E6 gene of HPVs 14, 21,
Delius"	and 25 can enhance the induction of anchorage independent growth of
200. .706	3Y1 cells by the HPV16 E7 gene, although again less effectively
/gene="E6"	than that of HPVs 5, 8, and 47.
200. .706	
/gene="E6"	
/note="putative"	
/codon_start=1	
/product="transforming protein E6"	
/protein_id="AA79394.1"	
/db_xref="GI:1020171"	
/translation="MADSTDSADGSPKRRHLEENNTSFLPEPLPATIRDLANIL	
EIPDDCIAPCNFCGNFLTHLECEDEKSLILMKDKHCVACCCVCAATATYEYN	
FYEYIVGDRIDEITGKSPFDIVACYMKMFLDSIEKLDICGRKFFPHKVGSKGI	
CRICKHPO"	
706. .1011	
/gene="E7"	
706. .1011	
/gene="E7"	
/note="putative"	
/codon_start=1	
/product="transforming protein E7"	
/protein_id="AA79395.1"	
/db_xref="GI:1020172"	
/translation="MIGKEVTLQDVLVLELNLQEPVQVDFLCEELPSQOETEEEL	
PERYAVKVVTPCGCKVKRLIFVNATGFAIRTFQNLLEELQCLCPGCRGCKRGKS"	
gene	998. .2809
CDS	/gene="E1"
	998. .2809
	/gene="E1"
	/note="putative"
	/codon_start=1
	/product="replication protein E1"
	/protein_id="AA79396.1"
	/db_xref="GI:1020173"
	/translation="MADPKGTSKGLDEDMCIVEAGCSQDVENDLELPRDPTDISE
	LDDNDLEQNGRELFLHOESKESELOLKRKYLSKRAVAQLSPRESITLSPQOK
	SKRLFEADDSGLECTLTFNEEDSVSEVPALDSQVPAEAGTDTYKELRASN
	KILMAKKEFEFGVPPNDLTROFKSKTCQCAWVSVAAYHDLLESSKOLLQDCD
	IVIRIGAMSLFLQCFQKNGRYHKLMTALNVHEKQIISEPKLRNVAAALFWK
	GAMSGATFPGYPPMIAOQTVGHSTVESAFAFDSAVQAFNNYLDLADIAVOYA
	KLAPEDSNVAVLANNOARYREVAASVPRKNCILLYGPNTGSAFTMSLIVLRG
	WSTIAKFLRYQOVNIMFLAALKDLHSVPKRCILLYGPNTGSAFTMSLIVLRG
	RVLSFVNSKQFQWLPQMSCKLALIDVTDPCWIMDTYILRNGLDGHVSLDCKHAP
	MOTKFPALLTSTINIVNHEVNYRYLHSHRTKGEFPNPPMKADMTPEFLTDQSKSF
	FRIMNOLLEISQDEDEGENGESQRFQCSANENHL"
	2751. .4262
	/gene="E2"
	2751. .4262
	/gene="E2"
	/note="putative"
	/codon_start=1
	/product="regulatory protein E2"
	/protein_id="AA79397.1"
	/db_xref="GI:1020174"
	/translation="MENLSDRFNVLDQLMNIYESAANTIESQTEHMQTLRKEVILY
	FAKQGVIRLYOVVPLAVSESAKQALGMALQLOSLQSEVAKESMLVDSATF
	RSPEENRHKKQVSEVLYDNDKNANAYTMARVYVYDDDDHKKSPGYNHGYE
	MOGFERYHYLFADDAKRYSTKGMEVNVKEEYAEVATDSTPDGSGADASTST
	TPATTDSTSLSSRKOSQOONTNKGRRYRRTTQTPORRRSSKSR
	SRSRRSRSRSRSRSYSSRSOSDOOYRRSGCOVSLTTATTTTATNTNSTOS
	GRGSSSTSSSKRRPRRGATIGSSGGRGRSSTSPSSKRGSESVRGRGISP
	DVQKSLQSDVSTRNTGRILDLALDPVTVLVGEPNTLKCFENRAKILVAGLYKAF
	STAMSVAGDSTETLGRMLISFSEQRKDFKTVKYGVDRSYSPDSI"
	3037. .4017
	/gene="E4"
	<3037. .4017
	/gene="E4"
	/note="putative"
	/codon_start=1
	/protein_id="AA79398.1"
	/db_xref="GI:1020175"
	/translation="IPVQRHLEALKLIISKQGCQRLFMITITKTMILPGCDMFT
	WMATTNGCIKVOAVSTQAYILCKELDTTFYLLMMOYDIAELDICKITLIRLCIL
	SPAPHPPTHQDKOTPPPPPPPPPTLPPSPRPENSNKTPKEEGTDGKPYG
	GSEPKRIKGDGRGSRGGRGKRGSDPPDPGPIPGGLRLTISRNTDSDPEKQ
	PSSLPPLPPPPPOPTTPBGQGRGHPPPPPPPNGHDGEGPLEAGVGGDDHPPP
	PAPNGHEESQSLGNVASTLTWTWESLFNQLVQEIQVLEDYTKLSIQ"
	4351. .5913
	/gene="L2"
	4351. .5913
	/gene="L2"
	/note="putative"
	/codon_start=1
	/product="minor capsid protein L2"
	/protein_id="AA79399.1"
	/db_xref="GI:1020176"
	/translation="MARAKVKRDSATNIRTCQKAGTQCPDVINKVESTYIADKILQ
	YGSAGVEFGIGTSTGKGTGCTGYVPLDEGPAVRYGNAPTYIRALVDTTGPSSIT
	PVDTLNDPEPTSSIYPLDSDPDLPGVEETIAETIHGPTPRPPDVAVTTSNOST
	AVLEVAPEPTPSKRVNTRQYHNSFQVITESTPTTGSSSLADHLIVTSGGQVIG
	GSTPELLELDQPPSKRYSEFELPEPTPRKSTPQRIQNTIIRRGGLTNRRILVOQNV
	ENPLFVSRSLVOFQNDPAEEVEYQTFEODIDTFNPPDRFDLITLGRPOYSE
	TPAGYKVSRLKRGFTIRRSQIGSOVHFYRDLSTINTEPDLIQLEHSGGATV
	VQGPVESTIDIVNDENPLSEDSAHSDLLDENEDPSQIVVGGRRSSSTVP
	REFTTSGSYVQDDTKGYVNAVPEEDQSTDTITITPPDLPVYIITFTDTSQFYLHPS
	LSRKFKRRKRYL"
	5929. .7485
	/gene="L1"

```
CDS
5929..7485
/ gene="L1"
/ note="putative"
/ codon_start=1
/ product="major capsid protein L1"
/ protein_id="AA09400.1"
/ db_xref="GI:1020177"
/ translation="MAVWAASGVLYLPSTPVARVQSTDEVYQRTNLYHYASDRLL
TNGHPYENVYVNSAKIKVPRVSGNHREVLKLPDNPRLADMSYNPDKERLWVA
CRGLEIGGQPLGVSGVSHPLFNKVGDPENDSSYKTPNSTDDRONVSPDKLOMEI
IGCAPCIGEHMDKAIPCATIDPFPSPGSPPELINSALQDCMDADIGYNLFEALQON
RSDVSLDIVNETCKYPOFLKQNDVYGDSCFTARKDCTARHFPVAGKGTGDIAG
QIDEGSMKNAYIIPMDQAOYKIGNSMYEPYVSGLSVSSDAQLFNRPVLOKQGN
NGICWFNOLFVTVNNTNFSSISVNPENDVSKIEENYKAESFOEYLRIHVEYELSL
TLOLCVPLTAEVLAQITAMANAATLEEMQLGFAPADNPIDHYRYIDSATKCPKPN
PKKEREDPKMKMKFMDVLTFRLLSLDDQYSLGRKFLFGAQLQOTTVNGTKTSSRVS
TKGIRKRRKN"

BASE COUNT      2426 a      1518 c      1680 g      2155 t
ORIGIN
Query Match      60.0%: Score 15: DB 59: Length 7779:
Best Local Similarity 100.0%: Pred. No. 1.4e+02:
Matches 15: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

OY      3 atctgcctgttgaa 17
      |||
Db      941 ATCTGCTGTTGAAG 955

RESULT 42
AF066865/c      8219 bp      DNA      PHG      26-NOV-1999
LOCUS      AF066865
DEFINITION      Bacteriophage TPW22 holin, lysin, integrase (int), and repressor
      protein (rap) genes, complete cds; atlp site; and unknown genes.
ACCESSION      AF066865
VERSION      AF066865.1 GI:6465902
KEYWORDS
SOURCE      bacteriophage TPW22.
ORGANISM      bacteriophage TPW22.
REFERENCE      1 (bases 1 to 8219)
AUTHORS      Petersen, A., Josephsen, J. and Johnsen, M.G.
TITLE      TPW22, a lactococcal temperate phage with a site-specific integrase
      closely related to Streptococcus thermophilus phage integrases
JOURNAL      J. Bacteriol. 181 (22), 7034-7042 (1999)
MEDLINE      20026817
PubMed      10559170
REFERENCE      2 (bases 1 to 8219)
AUTHORS      Petersen, A., Josephsen, J. and Johnsen, M.G.
TITLE      A novel lactococcal bacteriophage repressor located within a
      chimeric lysogenic region of the TPW22 genome
JOURNAL      Unpublished
MEDLINE
PubMed
REFERENCE      3 (bases 1 to 8219)
AUTHORS      Petersen, A.
TITLE      Direct Submission
JOURNAL      Submitted (18-MAY-1998) Department of Dairy and Food Science, The
      Royal Veterinary and Agricultural University, Copenhagen 30,
      Frederiksberg DK-1958, Denmark
FEATURES
source
      1..8219
      /organism="Bacteriophage TPW22"
      /db_xref="taxon:79219"
      <1..117
      /note="orfC"
      /codon_start=1
      /transl_table=11
      /evidence=not_experimental
      /product="unknown"
      /protein_id="AA012703.1"
      /db_xref="GI:6465903"
      /translation="SVVTNSSSSSGYTGTONIGSAVIGSTTIYWKRTA"
      141..407
      /function="formation of non-specific lesions in the
      cytoplasmic membrane"
      /note="orfB"
      /codon_start=1
      /transl_table=11
      /product="holin"
      /protein_id="AA012704.1"
      /db_xref="GI:6465904"
      /translation="MNOINMKLRKSKAPVALLPFLILIOIGASFGYKNNFVILN
      QOLAAVVAAPALLAIVGAVADPTTSLGSDRYLNKDKSEBK"
      404..1693
      /function="degradation of host cell wall"
      /note="orfA"
      /codon_start=1
      /transl_table=11
      /evidence=not_experimental
      /product="lysin"
      /protein_id="AA012705.1"
      /db_xref="GI:6465905"
      /translation="MKRLIKKRAIGMFAFVVAASGPAFAVAGDQVMSKNGTYGN
      EGYARDKAFSQIGTGTGTFVDQATYETQVASAIAOGKRAHYIWRVGSQSEVKA
      ALDRYLPKTIQPKNSIVALDYEGASGKQKQNPALILYGMRYKAAGTFPMYSYRY
      TLAVNYKQIILKEPKSLIAIAPYEVTPVPNTSFPSPMDGISVPGFTSYVAGGD
      GNVDELGTIDNGYKQGOEVKPNTPATPAIKNGEAMVEVKGNDVAGMTKAVNGAKH
      YATGETIPQWYKGPCHKIIRKNGETVILDGIMSLSVHDEVTIDASTOPTPAKSTI
      VKQDPLSGIASNMGCTNMOELARONSLNPNTIYTGQVIRFTGGSGSTAAATYTVRSG
      DNLSIASRLCTTYSLSVSMGINSPLIYVQTLNR"
      1695..1727
      /evidence=not_experimental
      1722..1739
      /rpl_family="r6"
      /rpl_type=direct
      1748..1755
      /rpl_family="r7"
      /rpl_type=direct
      1759..1766
      /rpl_family="r7"
      /rpl_type=direct
      1796..1826
      /evidence=not_experimental
      1820..1837
      /rpl_family="r6"
      /rpl_type=direct
      1841..1848
      /rpl_family="r5"
      /rpl_type=direct
      1857..1864
      /rpl_family="r5"
      /rpl_type=direct
      1917..1942
      /evidence=not_experimental
      1940..1947
      /rpl_family="r4"
      /rpl_type=direct
      1943..1956
      /rpl_family="r1"
      /rpl_type=direct
      1950..1956
      /rpl_family="r3"
      /rpl_type=direct
      1954..1967
      /rpl_family="r1"
      /rpl_type=direct
      1961..1967
      /rpl_family="r3"
      /rpl_type=direct
      1984..1991
      /rpl_family="r4"
      /rpl_type=direct
      complement(2001..2014)
      /note="atlp attachment site core region"
      2041..2053
      /rpl_family="r2"
```

```

repeat_region /rpt_type=inverted
2041. 2047 /rpt_family="r3"
repeat_region /rpt_type=inverted
2062. 2074 /rpt_family="r2"
repeat_region /rpt_type=inverted
2068. 2074 /rpt_family="r3"
repeat_region /rpt_type=inverted
2112. 2125 /rpt_family="r1"
gene /rpt_type=direct
complement(2137. .3204)
/ gene="int"
CDS complement(2137. .3204)
/ gene="int"
/function="involved in site-specific phage integration"
/codon_start=1
/transl_table=1
/evidence=experimental
/product="integrase"
/protein_id="AAFI2706.1"
/db_xref="GI:6465906"
/translation="MANFRKRGKTWQFLRSLKDNNGEYKKFEKGGYKTRKEAEAAADF
AKRLNHSFENDISLDYDFEKKMAVYKRPVTEATWYTRTLNLIDXTKPKPLAE
ETTPYQVALLNMSLLYRQESLDKPYFOIKSAMIAHEKVISNFPADPFAKASKILA
ARPEEYVYLADEYKLILAIAEKMRYTSFACYILAVTGMFAELGLTSTPMFRAKSKLA
KEISITQWYSITNNFAETRNKSSKRIPISSKIKLKIKKRYTHHNNKYDVPDK
LNNNGINKTKIVTAGRKRVHPSLRHSFASYLTKGIDLTIVSKLIGHBNLVNTLKYVA
HOLKEQNDNDYIRKTRPKL"
complement(3242. .3247)
-10_signal /evidence=experimental
-35_signal complement(3264. .3269)
CDS /evidence=not_experimental
complement(3331. .3948)
/ note="orf2"
/codon_start=1
/transl_table=1
/evidence=not_experimental
/product="unknown"
/protein_id="AAFI2707.1"
/db_xref="GI:6465907"
/translation="MTYIDELKEKQLDDGLITDPEFSAOKSHLESKEIYNADPT
SNSIEEGEKEVEYFDPKKNKCSGAKGTWITFRKLDNHFICANCTIKKGLTTSN
OETIKONTLDITFSDHDKTKSSOPTIKRIKAPVDNRAIKCPCKSTDVFMQGGK
GFSVSGAAGVALVGTGLAGFAGKKGKKOMHNCNGRVETKK"
complement(3989. .4177)
/ note="orf3"
/codon_start=1
/transl_table=1
/evidence=not_experimental
/product="unknown"
/protein_id="AAFI2708.1"
/db_xref="GI:6465908"
/translation="MRKPSLAKSFKAPTTIKYKRYKKALLIPYGGKGMKANPKKA
AYNKYKTFPSIMDLK"
complement(4201. .4764)
/ note="orf4"
/codon_start=1
/transl_table=1
/evidence=not_experimental
/product="unknown"
/protein_id="AAFI2709.1"
/db_xref="GI:6465909"
/translation="MLKQETIEOTIIEIENHGIDVIGDNSFPDIAITNNRKITIIYNP
QIAPFPLTHELHILINCIDIRNDYDSTSPCKRANTPAILKLNNEFPOGGTTEEL
YQIFLEVCGPEPKLTKIIVLKSITKSWDKPEVQHVTYHLDSTDDEPESNNYVSINDAC
HIDKMSLWSTLIDNLSKFNQSVI"
complement(4777. .5121)
/ gene="rap"
CDS complement(4777. .5121)
/ gene="rap"

```

```

/codon_start=1
/transl_table=11
/evidence=experimental
/product="repressor protein"
/protein_id="AA12710.1"
/db_xref="GI:6465910"
/translation="MDLYEKIKELASQKNVSIROYEERKIGFANGTIROMKKNPKINIK
VKDVAKYFNVSVDFTLGDNDQRRKEPVLDLADFPDDKKVWMDENVSDGKPLSDRYKN
AMKILGRLED"
5181. 5186
/evidence=not_experimental
5204. 5209
/evidence=not_experimental

-35_signal
-10_signal

Query Match
Best Local Similarity 100.0%; Prod. No. 1.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 gtttgaagcgagca 24
|||||
Db 5925 GTTTGAAGCGAGCA 5911

RESULT 43
AE004129 10881 bp DNA BCT 31-JUL-2000
LOCUS Vibrio cholerae chromosome I, section 37 of 251 of the complete
DEFINITION chromosome.
ACCESSION AE004129 AE003852
VERSION AE004129.1 GI:9654836
KEYWORDS
SOURCE
ORGANISM
Vibrio cholerae.
Vibrio cholerae
Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
1 (bases 1 to 10881)
Heidelberg,J.F., Eisen,J.A., Nelson,W.C., Clayton,R.A., Gwin,M.L.,
Dodson,R.J., Haft,D.H., Hickey,E.K., Peterson,J.D., Umayam,L.A.,
Gill,S.R., Nelson,K.E., Read,T.D., Tettelin,H., Richardson,D.,
Ermoлаeva,M.D., Vamathevan,J., Bass,S., Qin,H., Dragol,I.,
Sellers,P., McDonald,L., Uitterback,T., Fleischmann,R.D.,
Niernan,W.C., White,O., Salzberg,S.L., Smith,H.O., Colwell,R.R.,
Mekalanos,J.J., Venter,J.C. and Fraser,C.M.
Direct Submission
Submitted (14-JUN-2000) The Institute for Genomic Research, 9712
Medical Center Dr. Rockville, MD 20850, USA
Location/Qualifiers
1..10881
/organism="Vibrio cholerae"
/strain="N16961"
/serotype="O1"
/db_xref="taxon:666"
/chromosome="I"
/note="Biotype: El Tor"
115..1158
/gene="VC0415"
115..1158
/gene="VC0415"
/note="similar to SP:PI3519 GB:M22055 PRD:606191
PTD:1128967 GB:U00096; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11

```

```

/produet="rod shape-determining protein Mrec"
/protein_id="AAF93588.1"
/db_xref="GI:9654837"
/translation="MKRLGPMFNSNDLSIDLTANTLIYVGGQIVLDEPSVVAIROD
KRGKRTVAANAHAKOMLGRPCNITSIRPMKGVIADEPMAAIGALRIARSEPIS
VLRSPRIIVCVPGSTOVERRAIRSALGAGAEVLIIDEPMAAIGALRIARSEPIS
SMVIDIGGTAEVAVISLVGVYSSVIRIGDREDAIINVRNRSGLIGERAEKI
KHEIGASPGDVQVEIWEVRGNLAEGVRPSTLSNELLLEALDEPLTIVASVAVALE
OCPPELASDISENGWVLTGGALLKLDLRLMEETGIPVIADPLTCVARGGKALE
MDMHGDLFSEE"
1215..2105
/gene="VC0416"
1215..2105
/gene="VC0416"
/notes="similar to GB:M31792 SP:P16926 GB:M22055 PID:146818
PID:146888; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="rod shape-determining protein Mrec"
/protein_id="AAF93589.1"
/db_xref="GI:9654838"
/translation="MKPIRGSPISQILRLFEAVMLASLMLADSRIGAFSHVRELLNS
LVARIOYANDIRPVLFDDPFDYSISQOLRENNOLLKREVLTKSNLLILROYKEENQ
RIKRLIGSSFLNDEKKVYEVAAVDTSPTRHGVYIDKGRVGVGEGPITINKEGIVGO
VTFVAHNSRVLLIDPNTAIFVQNRNDIRVJASNGCOTQDIOLEHITPSTDIEVD
MVTSGIGVYEGYEPVAIVSQVDKTRREFASIKADPVEHFRLLYLILPWEDRL
OKYQADPQVLLLEENAAQ"
2092..2580
/gene="VC0417"
2092..2580
/gene="VC0417"
/notes="similar to GB:M31792 SP:P16927 PID:146819
PID:606189 GB:000096; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="rod shape-determining protein Mrec"
/protein_id="AAF93590.1"
/db_xref="GI:9654839"
/translation="MLSSDIKGRVILASFLALILQITPMPGTLDLFRPPULLLVCC
YVWIALPHRVNVSALVGLMLDLIGSTLGRGMMSSIIYLVANFLVLENNMLMO
OAVIYGLSMGLVLIHFCRYLIQNTVFNPLISVSGVIAIIPWMLFLIRRRHH
VK"
2523..3140
/gene="VC0418"
2523..3140
/gene="VC0418"
/notes="similar to GB:L08793 SP:Q02169 GB:M96343 PID:142853
PID:143161; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="maf protein"
/protein_id="AAF93591.1"
/db_xref="GI:9654840"
/translation="MAVDVETAPSTPLACEMTISKVLASGSPRRRELLAQMGYOF
EVVAVNYEERKAASPAOYVERLSRDKALAGALVAAEAVIGSDTIYVKDOVLEK
PROFADAKRMILKLSQSOHOVMTCVSTCRGITHSVYTVTEWFKTISQELIAYWQS
GEPDCKAGSTIGLIGRFRYTRIGSHAVVGLPLYETDOLLKF"
3166..4635
/gene="VC0419"
3166..4635
/gene="VC0419"
/notes="similar to SP:P25537 PID:48826 PID:606187 GB:000096
PID:1789645; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="cytoplasmic axial filament protein"
/protein_id="AAF93592.1"
/db_xref="GI:9654841"
/translation="MSAEILLNVPSETRVAMIEGILQEIHTIEREAKRGIVAGNIYK
KYSRVLPGQAAFDVIGDKAAFLHASDIYVTECVASEKQOQVARDISIELROGD
IVQVVKDPLGKARLITDITLPSRTLVFPPGASHVGSORISEEAREERLIKIVSH
YCDHGEFLITTAAGANEKELTDAAFLKRLMSKVIERSKYNKRSMLTGEIGLAOR

```

```

ILRDFVGTETLKIILDSROEFENLKEFTSEYVELTAKLELYHGDKPIEDMYOTENEI
ORSLERKESGGVLLIDQTEAMTIDITNGAFVGRNLEETIFMNIENAEITAKIRO
LRLRLGIIILIDPMDASDEHRRKRVLTSLIDAKMRVNTINGTOLGIVEMETKR
TRESTEHLCTCPTCEBRSQSVKVENYCEPILREITRVNRAVDADKPFVYASPAVAD
TLQGESHALALEVFIGEVRIOAEPLIYNODEPVM"
4648..8523
/gene="VC0420"
4648..8523
/gene="VC0420"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF93593.1"
/db_xref="GI:9654842"
/translation="MISTVTRGRTIMLLVILITVLAATVATLRILLPQNRFOAI
QIHLNONSVOVALADVOGWNTPTSPISLOTLOAHNPPDSNDIOLNASVETIEDILO
SLMORPVVADLTNGVLYDLRAIDMLAEONPNRSGRGVYKQDLILLRDLDF
TLKNSAIIYRTFAGDLRODIEKLRMQNGLRHAEVYSIAGINISLVSAFIDH
GSLRVSDEYVSADKRVYLPMLTRVYLDKQDQIGQGVLSANAMTLEHNDKDYVE
KSEIYNONGEOTHELLESGIVELKPTKEGMVNAHSLRLSDEDELPILVAMQ
PDEMRNLQMLINELLPLAKLIPESOTLNMHTLTKRGTLEDVRIAKGATLESIRY
SASLIDGIAQWELLPOVNAQOJOGSPKALIKASWDDVLYGEVFOAPLNIROG
AVNLVMOQOYQMSLMSDVYVATPDIALGAFRLDPDDPYPYRHNGVFOAGLAE
TWRYLPTLALGRELIDYSTAIOAOVNTAKLIMGALDOEPYRHNHNGVFOAGLAE
AKFAPDTAMPPTIDLDLLENDAMYLDSRSATLMEVSAERTIGRIPELAEGLIET
EYASAOQCNALIRNYMATPLVDSGALTLTOVAGPVSAGDLARLLKOPVSIDFGEDAKRG
FATLSNNADIDPMPSLTSSGKIIFENDRVSAGDLARLLKOPVSIDFGEDAKRG
YAGIDVMDGMEYKPLIFPYGGRWLSRYKGAHPMOASDIOINDGFFYLODGAADR
GLERSYFPLPKALKVYGQALIQASGNDQMSARQLQOAKOATDITLTPVPLAKT
NLVLGQSEKISFHVGHVHVLRSQAFENLNDQILSTINEBPKRNSKSLASNTAPFM
PERVDAVKEITSPAGLMDVLDVLMARKKOLMLNDLSQETIKQANVIEYDLSALE
RLHLEPPLQAGEALLVDLDKOKPLITDFPKRHOWMPNLTITIKQFVLOGKTIQO
ANDPOROGDTLMLKSIDPESGNOLHVTGTWTDQSRQOMNDMDKGNNSDLMAR
PGTNSGTOARPEITASTQMDGAPMSQVNTLOGVDYRKGKGVISDVSGARLLGLF
SLDSITRKMOJLDSYVFDKGMADSISSGSELISGIPVNNIKMDAVAGEMTIKGLAD
LNRITDAEVENFVDITSGIPVLTAFVTPOTALVYLAITYISVVEVFQVNEVK
GPLDSPVKELSLRSKGEFKLEKRLTE"
8532..9359
/gene="VC0421"
8532..9359
/gene="VC0421"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF93594.1"
/db_xref="GI:9654843"
/translation="MORVGLVQMTSGPEISRNLTLYKQVANKLAQOAGOMITPENAL
LIGNREOYHOQAESLDHGPVQHALASLAKRYCWMILIGSMPIRHAGCTTSLMNAO
GERVAVYDKLHMFVDVADGHORYRESERTPEQOVVYSTPFGALGISTCYDVRPH
LYADLRQGAOILLVPAFTAVTGOAHWEVLARARIEQCWVIAVGQGHHPGRET
WGHSMVISPWGEEVIANGAEVQSKVEFDLATLDSVRAMPITQHRTFHOLEMTH"
9359..10804
/gene="VC0422"
9359..10804
/gene="VC0422"
/notes="similar to SP:P46473 PID:606183 GB:000096
PID:1732437 PID:1789640; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="cltd protein"
/protein_id="AAF93595.1"
/db_xref="GI:9654844"
/translation="MTMNOIETALLAPAGLTEDDIADTLSSIATROIDVADYFGSSW
HESIYLERISITIKDPSFNDGCVRAVSGEYTGAFSDQJOLBGLKQSAIANGIAQO
GONTVQAFKRAOHPAFYGAHNPLESWEKQOKTELKOLDAYIRTEKPLTTEVSYSLS
GVHEOILVADTDGTAGDTRPVLRSLSVLAOKDRRERSAGGGGFGFYDFOSEA
GQKIAESFADATIRQALVNLVAAPAGTAPVYIGSSMPVYLIAHNGHLEDDPNK
GSSVAGVYGRKVTSDICTYVDDGTLDKRLSLVNDDEGSGQIVNVIIEGKIAGYO
DKLNARLNGVAPTGNGRESYVAHLPMPRMTNVTMLPEHPEETIASVKKGIAPNFG
GGQVDTSGKFVFSASEAVYLTENCKITRPKIGATLIGSGIEAMQOVSMDNALDRG

```

```

BASE COUNT      2677 a 2487 c 2982 g 2735 t
ORIGIN
          GVCGKAGCAGSVGVGQFPLKDALTVGGTIE"

Query Match      60.0%; Score 15; DB 1; Length 10881;
Best Local Similarity 100.0%; Pred. No. 1,4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      2 gacgcctgtttgaa 16
        |||
Db      6397 GATCGTGTGTTGAA 6411

RESULT 44
AE006532      11062 bp      DNA      RCT      13-APR-2001
LOCUS      Streptococcus pyogenes strain SF370 serotype M1, section 61 of 167
DEFINITION      of the complete genome.
ACCESSION      AE006532 AE004092
VERSION      AE006532.1 GI:13621984
KEYWORDS
SOURCE      Streptococcus pyogenes.
ORGANISM      Streptococcus pyogenes.
REFERENCE      1 (bases 1 to 11062)
AUTHORS      Ferrerelli,J.J., McShan,W.M., Adijic,D., Savic,D., Savic,G., Lyon,K.,
              Primeaux,C., Sezate,S.S., Surorov,A.N., Kenton,S., Lal,H., Lin,S.,
              Qian,Y., Jia,H.G., Najjar,F.Z., Ren,Q., Zhu,H., Song,L., White,J.,
              Yuan,X., Clifton,S.W., Roe,B.A. and McLaughlin,R.E.
              Complete genome sequence of an M1 strain of Streptococcus pyogenes
              Proc. Natl. Acad. Sci. U.S.A. 98 (8), 4658-4663 (2001)
              21192684
              MEDLINE
              PUBMED
              11296296
              2 (bases 1 to 11062)
              Ferrerelli,J.J., McShan,W.M., Adijic,D., Savic,D., Savic,G., Lyon,K.,
              Primeaux,C., Sezate,S.S., Surorov,A.N., Kenton,S., Lal,H., Lin,S.,
              Qian,Y., Jia,H.G., Najjar,F.Z., Ren,Q., Zhu,H., Song,L., White,J.,
              Yuan,X., Clifton,S.W., Roe,B.A. and McLaughlin,R.E.
              Direct Submission
              Submitted (10-APR-2001) Department of Microbiology and Immunology,
              University of Oklahoma Health Sciences Center, 940 SL Young Blvd,
              Oklahoma City, OK 73104, USA
              Location/Qualifiers
              1. 11062
              /organism="Streptococcus pyogenes"
              /strain="SF370"
              /serotype="M1"
              /db_xref="taxon:1314"
              85. 1248
              /gene="SPY0808"
              85. 1248
              /gene="SPY0808"
              /note="Best Blastp hit = sp|P39587|YMBD_BACSU HYPOTHETICAL
              44.4 KD PROTEIN IN EPR-GALK INTERGENIC REGION
              >g1|629026|pir|1|S39674 ymbd protein - Bacillus subtilis
              >g1|413943|emb|CA51575.1| (X73124) lpa-19d [Bacillus
              subtilis] >g1|263637|emb|CAB15862.1| (Z99123) alternate
              gene name: lpa-19d
              similar to hypothetical proteins [Bacillus subtilis]"
              /codon_start=1
              /transl_table=11
              /product="conserved hypothetical protein"
              /protein_id="AAK33745.1"
              /db_xref="GI:13621985"
              /translation="MNLKLYDSFVEKKLAGVOLLDEKFSNINERKQVLYVTKSNR
              PIGTAVLSKQKNGIGWYLGPEKIDLSISYVSEVAKARQDFAQSDETNAVLFNQ
              EGDGEGVITDLYGFAVFSWNAFVSKEMIEAOVPEVKGAEVCECRGPDPT
              ETALHLELAQETFSILENGIAYOVFNEGIMGIFLDQHDVRAIYDGLAMGSLIN
              LESTYAFSAVAAAGALLETTSVDAKRSRSLAHFEHMLNLASHHFVVMVFET
              KYARKKLIPIVIVIPPSFARKKOTFSVRDHLKLTALDLILSPKGIITASTNA
              NMTVSQRFKQIIRKIGFSRRPESMTLQOLPSDFITNKADERSNYLKVFTIVRE"

gene      1245. 1931
          /gene="arod"
          /note="SPY0809"
          1245. 1931
          /gene="arod"
          /BC_number="4.2.1.10"
          /note="Best Blastp hit = sp|P24670|AROD_SALTI
          3-DEHYDROQUINATE DEHYDRATASE (3-DEHYDROQUINASE) (TYPE I
          DHOASE) >g1|96689|pir|1|S15652 3-dehydroquininate dehydratase
          (BC 4.2.1.10) arod - Salmonella typhi
          >g1|7546223|pdb|1OFE|A Chain A, The Structure of Type I
          3-dehydroquininate dehydratase from Salmonella typhi
          >g1|7546224|pdb|1OFE|B Chain B, The Structure of Type I
          3-dehydroquininate dehydratase from Salmonella typhi
          >g1|47642|emb|CA38418.1| (X54546) 5-dehydroquininate
          hydrolyase (3-dehydroquinase) [Salmonella typhi]"
          /codon_start=1
          /transl_table=11
          /product="putative 3-dehydroquininate dehydratase"
          /protein_id="AAK33746.1"
          /db_xref="GI:13621986"
          /translation="MRIAIVPMRPHFDEAQAIDISKEDVNLIERADFLPKDEIVAV
          APAIEPKAGEKIEIFLRTVOEGGNITLSQEVYDIKEINAIYNPYIDREYTHRS
          VEOEMLDPEPNLILSYHNEETPENLMEAFSEMTLAPRVKAIYAVPQSDVDLDMNY
          TRGFRTLNPEDEFPATISMGKIGRLSRFAGDVIGSSWTIVSLDHVSGGYTLNDMKRI
          IEVLEMDISN"
          2025. 3191
          /gene="arof"
          /note="SPY0810"
          2025. 3191
          /gene="arof"
          /note="Best Blastp hit = db|BA005375.1| (AP001512)
          chorismate synthase [Bacillus halodurans]"
          /codon_start=1
          /transl_table=11
          /product="putative chorismate synthase"
          /protein_id="AAK33747.1"
          /db_xref="GI:13621987"
          /translation="MRYLTAGESHGPSLTAIIEGIPAGLTILHPADIDHELORRGGY
          KGARMSIETDRVQISGVRHGKITGATITLVINKDQKLDVMAVAKKIALEIETLKLR
          VKHPRGHADLVGKIKYHFNDRALERSRARETMMVANGAVAKKIALEIETLKLR
          ILIFGPTITIPSKISFRELORALHSELSTVPRKQEEIKTITDKEGDTGGIT
          ETIYQGPATIGSIVQDKKLDKALQAVLSINAFGVEFGAEFGKFGQSGVMDI
          TWTFQGYGRQTNHVGFEFGMTTQGPLVVGVMKPIPLTKPLMSVDIDSHRYKAT
          VERSDPTPALPAAGVIMENVATVLAKEILETFESTTMSLQKAFSDYRAVYQGF"
          3252. 3593
          /gene="SPY0811"
          3252. 3593
          /gene="SPY0811"
          /note="Best Blastp hit = pir|A69828 hypothetical protein
          yhea - Bacillus subtilis >g1|2226157|emb|CA47444.1|
          (Y14080) hypothetical protein [Bacillus subtilis]
          >g1|2633315|emb|CAB12819.1| (Z99109) yhea [Bacillus
          subtilis]"
          /codon_start=1
          /transl_table=11
          /product="conserved hypothetical protein"
          /protein_id="AAK33748.1"
          /db_xref="GI:13621988"
          /translation="MSQFIYDYANOLERAVRLPTEQKYLEKKAIOADVSAASELDE
          EVAMQKRIQGMOSGOMPTAREOTSIQELISQKIEANDIKAYFAEQALSYMSIER
          IVFAPLKLKLVK"
          3814. 5166
          /gene="gor"
          3814. 5166
          /gene="gor"
          /BC_number="1.6.4.2"
          /note="Best Blastp hit = db|BA076640.1| (AB019579)
          glutathione reductase (GR) [Streptococcus mutans]"
          /codon_start=1
          /transl_table=11
          /product="putative glutathione reductase (GR)"

```


Query Match	Score 15:	DB 1:	Length 11062:
best Local Similarity	100.0%;	Pred. No. 1.4e+02:	
			<p>/protein_id="AAK33749.1" /db_xref="GI:13621989" /translation="MVIPIYIVIGGSGAGIASNRRAMGAKVLAEGKEIGTCVN LGQGVKVMYGAQADILGTIVKADYDFDEKKEKDFKOLKANQAVITDIHASYERG FEQGVGRVIRYIAVFNDAHVEIAIGOLYTPHILATIGHHPVPEIDIAQAGISDGF FALDEPQVIRTAHVAKGYIAVELALVALGSKTIDPLIRHPRPLRSFKPTIVDYLVDFG AVNPDRHRTAEVAKGAYKNDDESLLTLKGQEEVDPOLIMAIRKPNLEGFSLDKG VTLNDKGIETDAENSTSYGIVAVGVNCKALATPAVAGRPLSRLEPNGKTDDEL DYQNAVITFHPVIGSVGISEEAQVAYQGEAKVTQSRPTSEMTATITHROPCLMK LVTGVDEKIVTGLHIGYGVDEMLIYKMGATKADPDNVIYAIHPGSEEFVYMR " complement(5254. . 6522) /gene="foic.2" /note="Spy0814" complement(5254. . 6522) /gene="foic.2" /note="Best Blastp hit = dbj BA06756.1 (AP001517) foyl1-polyglutamate synthetase [Bacillus halodurans]" /codon_start=1 /transl_table=1 /product="putative foyl1-polyglutamate synthetase" /protein_id="AAK33750.1" /db_xref="GI:13621990" /translation="MTTYDSISWTHFFRKANGKRTDLARMWLEALGRPODKPPAHV VPTGKSLSTVYLOHILISTSGVGTCTSPRTISPHRDIQSPISDLGAQVSYI RPIITKSLLETNMDRPTEFELVYTLIMCTYFANLKPVIDIIEAGISGKTATNTFAL AVCPDSISFDHQRGLYSIAEIAHQKAEVIRAKPEVITIGOLEASOVFOOVTKASS RLYOLGKDFLNPSEKTFSEFQHDNLNTRKLKLGQHKANACIALMTQOLRKSP KISRTKTOGKIEATWPGRSEFIOPNLLIDAHNPDISIAKSLLOEFPKPHILF AGLRKPLADLQOLDVEFEVSATNEDEPEALVLDYDPKRYQVQVDYDMLQGSVTSDE LFVYVGLSYFSEVROYCLSNANIFKNIF" complement(6552. . 6992) /gene="Spy0815" complement(6552. . 6992) /note="Best Blastp hit = gi 6323990.1 yrf1-6p [Saccharomyces cerevisiae] >gi 17306601 sp P538191 YNF9_YEAS_HYPOTHEICAL. 211.1 KRA PROTEIN IN COS1 5 REGION >gi 21327931 pit1 S63325 probable membrane protein YNL339c - yeast [Saccharomyces cerevisiae] >gi 2258153 emb CAA96276.1 (Z71615) there is a second 'v' element in the left: subtelomeric region of chromosome XIV: ORF YNL339c [Saccharomyces cerevisiae]" /codon_start=1 /transl_table=1 /product="hypothetical protein" /protein_id="AAK33751.1" /db_xref="GI:13621991" /translation="MSSRYSRQOKPRKSGVPTPHITGTPAFOKSIALIGSIITIV ATTISRALQVPPDKTSNDTPKRESSNTVIKIEKSSGQGNHGTSETTDNKTTLPE SSSTPSNHTTTPAETPSSATETGQIIPNATGADANTGLPTNP" 7227. . 8348 /gene="Spy0816" 7227. . 8348 /gene="Spy0816" /note="Best Blastp hit = sp P31672 NIFS_LACOE NIFS PROTEININ HOMOLOG >gi 198662 pir sl16047 nitrogenase cofactor synthetase protein nifs - lactobacillus delbrueckii >gi 43985 emb CAA43493.1 (X61190) nifs-like gene lactobacillus delbrueckii]" /codon_start=1 /transl_table=1 /product="putative iron-sulfur cofactor synthetase protein" /protein_id="AAK33752.1" /db_xref="GI:13621992" /translation="MIVFDNATTIPYGEALKTQYEVATKLYGNPSSLHOIGTASNR LEASKQIAGLIGKXSEIFPTSGCTEASNNALIGIAEKNAPKHHIISAIHPHAYS ESVKMLTIQGFEEVSTAPVTGQVDVYNALVETIRPDTLISIMAVNEMGAIQIPRAG SNLANOPTITFHVDAVQAIQKIPLCDMTNRYDLASFSGHFRHSVNRGAIQIPRAG RLNPLISQGGEOELRITTEVNAVGIASMAKRLIVTEKQVSVLPEKLTAMRQVIKSLSI APDVTVESAQEGEAFNLTIFGIRGVRGEVIVHAPEKYEIVITTSACSSKAGPAGS</p>

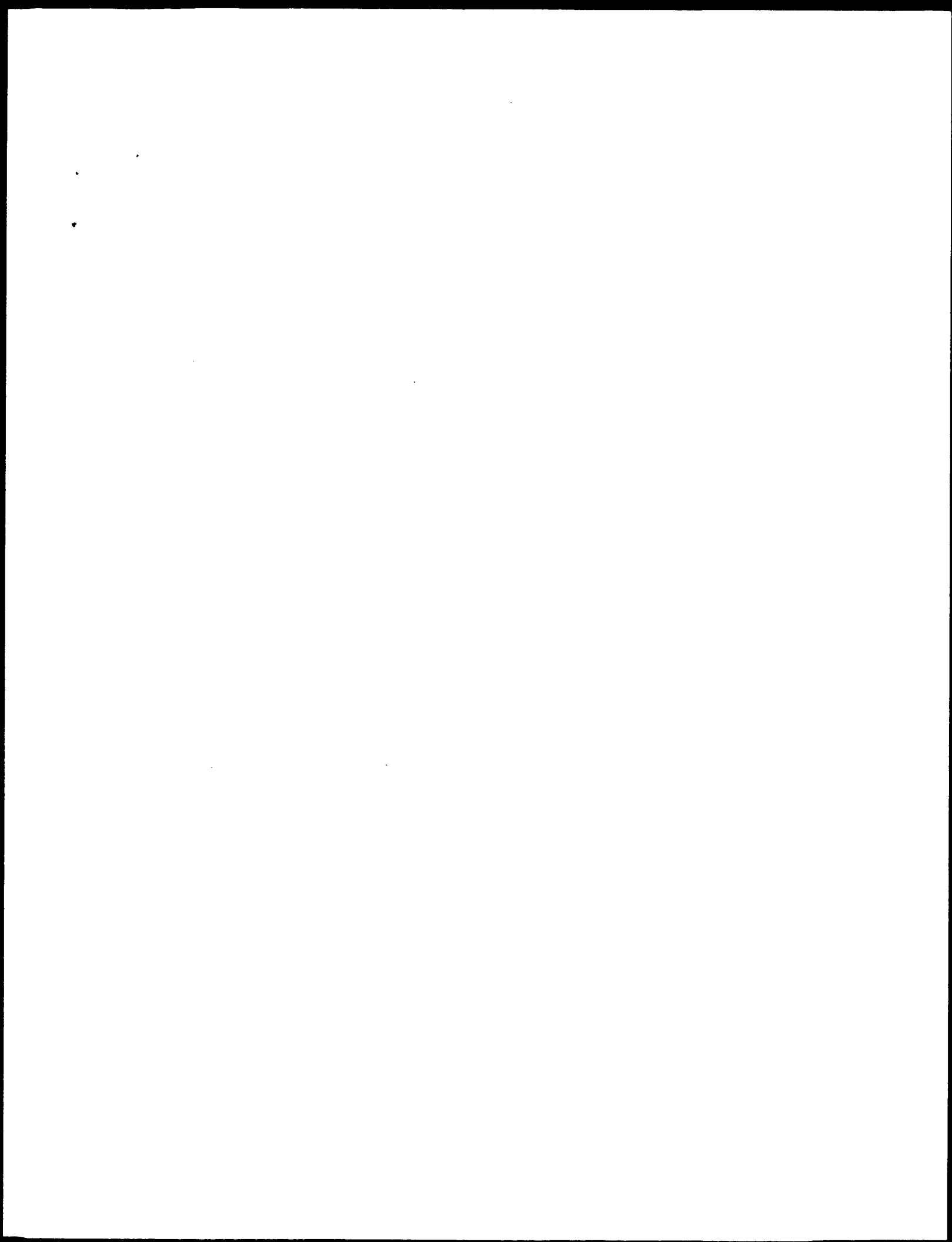
Matches	15:	Conservative	0:	Mismatches	0:	Indels	0:	Gaps	0:
QY	11	tttgaagcgcagcag	25						
Db	3510	TTTGAAAGCGCAGAG	3524						
RESULT	45								
AE004322/c									
LOCUS									
DEFINITION		Vibrio cholerae chromosome I, section 230 of 251 of the complete chromosome.							
ACCESSION		AE004322							
VERSION		AE003852							
KEYWORDS		AE004322.1 GI:9657104							
SOURCE									
ORGANISM		Vibrio cholerae.							
REFERENCE		Vibrio cholerae Bacteria: Proteobacteria; gamma subdivision: Vibrionaceae: Vibrio.							
AUTHORS		1 (bases 1 to 11328)							
		Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwin, M.L., Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Umayam, L., Gill, S.R., Nelson, K.E., Read, T.D., Tettelin, H., Richardson, D., Ermolaeva, M.D., Vamathevan, J., Bess, S., Qin, H., Dragol, I., Sellers, P., McDonald, L., Utterback, T., Fleischmann, R.D., Niernan, W.C. and White, O.							
TITLE		DNase sequence of both chromosomes of the cholera pathogen Vibrio cholerae							
JOURNAL		Nature 406 (6795), 477-483 (2000)							
MEDLINE		20406833							
REFERENCE		2 (bases 1 to 11328)							
AUTHORS		Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwin, M.L., Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Umayam, L.A., Gill, S.R., Nelson, K.E., Read, T.D., Tettelin, H., Richardson, D., Ermolaeva, M.D., Vamathevan, J., Bess, S., Qin, H., Dragol, I., Sellers, P., McDonald, L., Utterback, T., Fleischmann, R.D., Niernan, W.C., White, O., Salzberg, S.L., Smith, H.O., Colwell, R.R., Mekalanos, J.J., Venter, J.C. and Fraser, C.M.							
TITLE		Direct Submission							
JOURNAL		Submitted (14-JUN-2000) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA							
FEATURES		Location/Qualifiers							
source		1..11328							
		/organism="Vibrio cholerae"							
		/strain="N16961"							
		/serotype="O1"							
		/db_xref="taxon:666"							
		/chromosome="I"							
		/note="biotype: El Tor"							
		complement(95..286)							
		/gene="VC2521"							
		/note="identified by Glimmer2; putative"							
		/codon_start=1							
		/transl_table=1							
		/product="hypothetical protein"							
		/protein_id="AAF95663.1"							
		/db_xref="GI:9657105"							
		/translation="MAAFEDVLAIGIKREVEVPINYPPLALIPVCHSKRQPKLTPRLSSACRLCMKRFICFLDAK"							
		524..1522							
		/gene="VC2522"							
		524..1522							
		/note="identified by Glimmer2; putative"							
		/codon_start=1							
		/transl_table=1							
		/product="conserved hypothetical protein"							
		/protein_id="AAF95664.1"							
		/db_xref="GI:9657106"							
		/translation="MAFIFYVYGVNMLIEAVFLIGLAIYWGADKLTVGSAALRNIGISPLVIGMTILAKGSSARPMYNSATALEGGTPDPAVGVGSNTANIALIIGITRNIVKPLVSGSAVLRRELPMVIGVTLIAGAILMDNHLGFEGLILLAIAPFLVMAQVS							

Wed Oct 10 07:45:48 2001

us-09-396-196f-4.oli.rge

Page 39

Search completed: October 9, 2001, 15:45:19
Job time: 14705 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 15:52:08 : Search time 469.84 Seconds
(without alignments)
33.410 Million cell updates/sec

Title: US-09-396-196f-4

Perfect score: 25

Sequence: 1 ggcgtcgtgttgaagcgagcagcag 25

Scoring table:

OLIGO_NUC
Gapop 60.0, Gapept 60.0

Searched: 730101 seqs, 313950809 residues

Word size: 9

Total number of hits satisfying chosen parameters: 36461

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database:

1: /SIDSI/gcgdata/geneseq/NA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/NA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/NA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/NA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/NA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/NA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/NA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/NA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/NA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/NA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/NA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/NA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/NA1993.DAT.*
14: /SIDSI/gcgdata/geneseq/NA1994.DAT.*
15: /SIDSI/gcgdata/geneseq/NA1995.DAT.*
16: /SIDSI/gcgdata/geneseq/NA1996.DAT.*
17: /SIDSI/gcgdata/geneseq/NA1997.DAT.*
18: /SIDSI/gcgdata/geneseq/NA1998.DAT.*
19: /SIDSI/gcgdata/geneseq/NA1999.DAT.*
20: /SIDSI/gcgdata/geneseq/NA2000.DAT.*
21: /SIDSI/gcgdata/geneseq/NA2001.DAT.*
22: /SIDSI/gcgdata/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	25	100.0	1041	AA01303	E. coli biotin syn
2	25	100.0	1084	AA01329	E. coli Bio B gene
3	25	100.0	1121	AA06046	Sequence encoding
4	25	100.0	5872	AA062386	Biotin biosynthesi
5	17	64.0	5504	AA020559	Polynucleotide seq
6	16	64.0	924	AA071927	Corynebacterium q1
7	15	60.0	1689	AA023024	Rat kd312 polypept
8	15	60.0	2022	AA044132	Zea mays DNA fragm
9	15	60.0	3079	AA023025	Rat kd312 genomic
10	14	56.0	336	AA05487	Novel human polynu
11	14	56.0	379	AA07319	EST clone Bf797.

C 12	14	56.0	619	21	AAA44124	Human secreted exp
C 13	14	56.0	753	20	AA003347	Nucleotide sequenc
C 14	14	56.0	1173	20	AA003358	Nucleotide sequenc
C 15	14	56.0	1244	21	AA050896	Soybean PI metabol
C 16	14	56.0	1536	11	AA004684	SalI/kpnI fragment
C 17	14	56.0	1998	20	AA00367	Nucleotide sequenc
C 18	14	56.0	2087	20	AA00343	Nucleotide sequenc
C 19	14	56.0	2165	20	AA00379	Human fatty acid t
C 20	14	56.0	2405	22	AA027417	Human PRO703 nucle
C 21	14	56.0	2574	20	AA039871	Human PRO703 nucle
C 22	14	56.0	2574	21	AA078481	Human H-beta 58 fa
C 23	14	56.0	2574	21	AA058239	Human colon cancer
C 24	14	56.0	2669	20	AA025020	Staphylococcus aur
C 25	14	56.0	2675	21	AA098021	Sequence encoding
C 26	14	56.0	7778	18	AA074359	Sequence encoding
C 27	14	56.0	12284	11	AA006001	Hog cholera virus,
C 28	14	56.0	12284	15	AA071308	Genomic fragment
C 29	14	56.0	12284	17	AA029591	Septoria nodorum s
C 30	14	56.0	63563	22	AA028546	Septoria nodorum s
C 31	13	52.0	17	19	AA094360	Septoria nodorum s
C 32	13	52.0	17	19	AA062508	Septoria nodorum s
C 33	13	52.0	23	16	AA094361	Septoria nodorum s
C 34	13	52.0	23	19	AA062509	Septoria nodorum s
C 35	13	52.0	128	21	AA082189	Septoria nodorum s
C 36	13	52.0	157	21	AA012134	Septoria nodorum s
C 37	13	52.0	192	21	AA01863	Human secreted pro
C 38	13	52.0	251	19	AA012455	Human blallelic po
C 39	13	52.0	299	21	AA09435	Fusarium venenatum
C 40	13	52.0	300	21	AA001371	Human colon cancer
C 41	13	52.0	384	18	AA01888	Human ORK27 gene.
C 42	13	52.0	391	22	AA065138	Novel human polynu
C 43	13	52.0	398	22	AA066457	Novel human polynu
C 44	13	52.0	400	21	AA056847	Mycoplasma hyopneu
C 45	13	52.0	423	20	AA029975	Fragment of human

ALIGNMENTS

RESULT 1
ID AAX01303 standard; DNA: 1041 BP.
XX AAX01303;
DT 12-APR-1999 (first entry)
XX
DE E. coli biotin synthetase (Biot) coding sequence.
XX
KW DAP aminotransferase; diaminopelargonic acid; transgenic plant;
XX biotin synthase; biotin production; vitamin H; Biot; ss.
XX
OS Escherichia coli.
XX
XX US5669719-A.
XX
XX 09-FEB-1999.
XX
XX 30-APR-1997; 97US-0846338.
XX
XX 30-APR-1997; 97US-0846338.
XX
XX 08-MAR-1995; 95US-0401068.
XX
XX (NOVS) NOVARTIS FINANCE CORP.
XX
XX Patton DA;
XX
XX PI
XX WPI, 1999-152902/13.
XX
XX P-PDSB; AAW73906.
XX
XX Transgenic plants with high biotin levels - transformed with DNA
XX encoding di-aminopelargonic acid amino-transferase or biotin
XX synthase
XX

XX Example 2; Column 37-40; 34pp; English.

CC This sequence encodes the E. coli biotin synthetase (BIOB). The gene can
CC be used in the transgenic plant of the invention. The transgenic plant,
CC plant cell or plant tissue is transformed with a chimeric gene encoding
CC diaminopelargonic acid (DAP) aminotransferase or biotin synthase and
CC produces more biotin than a non-transgenic plant, cell or tissue. The
CC plant is used as an improved dietary source of biotin (vitamin H) for
CC humans or animals.

XX SQ Sequence 1041 BP; 262 A; 273 C; 305 G; 201 T; 0 other;

Query Match
Best Local Similarity 100.0%; Score 25; DB 20; Length 1041;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 ggatctgctgttgaagcgagcag 25
|||||
Db 63 ggatctgctgttgaagcgagcag 87

RESULT 2

AA091329
ID AA091329 standard; DNA: 1084 BP.

XX AC AA091329;

DT 15-FEB-1990 (first entry)

XX DE E. coli Bio B gene.

KW E. coli; Bio B gene; biotin.

XX OS Escherichia coli.

XX FH Key Location/Qualifiers
FT CDS 24..1064
/*tag=a

PN GB2216530-A.

XX PD 11-OCT-1989.

PF 17-MAR-1989; 89GB-0006210.

PR 22-MAR-1988; 88GB-0006804.

XX PR 17-MAR-1989; 89GB-0006210.

PA (UKAG-) UK MIN. AGRIC. FISH.

PI Pearson BM, McKee RA;

DR WPI; 1989-295085/41. P-PSDB P91392

XX Plasmid contg. gene(s) for expression of biotin synthetase enzymes
PT - derived from E. coli and capable of replication and expression in other
PT microorganisms, esp. yeast.

XX PS Table 3; page 33-4; 52pp; English.

CC The gene can be used in a plasmid for expression of enzymes of the biotin
CC synthetic pathway. Pref. control sequences for expression in S. cerevisiae
CC are plasmids pMA91, pMA36c, PKV49 and pCK495, and plasmid pCK965 for
CC Lactobacillus. Insertion of bio B improves biotin yields in
CC microorganisms which export biotin, or enables growth in media contg.
CC little or no biotin of organisms unable to synthesise biotin for their
CC own use.

XX SQ Sequence 1084 BP; 271 A; 286 C; 318 G; 209 T; 0 other;

Query Match
Best Local Similarity 100.0%; Score 25; DB 10; Length 1084;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ggatctgctgttgaagcgagcag 25
|||||
Db 86 ggatctgctgttgaagcgagcag 110

RESULT 3

AA060496
ID AA060496 standard; DNA: 1121 BP.

XX AC AA060496;

DT 17-OCT-1991 (first entry)

XX DE Sequence encoding biotin synthesising enzyme.

KW Biotin synthetic enzyme; E. coli; desthiobiotin; ds.

XX FH Key Location/Qualifiers
FT CDS 42..1082
/*tag= a

PN JP61149091-A.

XX PD 07-JUL-1986.

PF 24-DEC-1984; 84JP-0272605.

XX PR 24-DEC-1984; 84JP-0272605.

XX PA (NIPS) NIPPON SODA KK.

DR WPI; 1986-216622/33.

XX DR P-PSDB; AAF60536.

XX PT Double stranded DNA encoding biotin synthesising enzyme -
PT comprises transformed mutant E. coli strain contg. cyclic doubled
PT stranded DNA encoding biotin synthetic biotin enzyme as a plasmid.
XX PS Disclosure; Page 534; 23pp; Japanese.

CC The sequence may be expressed by a transformed E. coli host, cultured
CC in a medium containing desthiobiotin.

XX SQ Sequence 1121 BP; 290 A; 301 C; 319 G; 211 T; 0 other;

Query Match
Best Local Similarity 100.0%; Score 25; DB 7; Length 1121;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 ggatctgctgttgaagcgagcag 25
|||||
Db 104 ggatctgctgttgaagcgagcag 128

RESULT 4

AA062386
ID AA062386 standard; DNA: 5872 BP.

XX AC AA062386;

DT 16-NOV-1994 (first entry)

XX DE Biotin-biosynthesis genes contg. plasmid pB030A-15/9.

KW Biotin: expression; enterobacteria; vitamin H; synthesis;
KW plasmid; pB030A-15/9; bioB; bioC; bioD; bioA;
KW promoter; plac; biotin synthase; KAPA synthase;
KW 8-amino-7-oxononanoate synthase; pimeoyl-CoA; DTB synthase;

```

KM dethiobiotin synthase; DAPA synthase;
KM S-adenosyl-L-methionine: 8-amino-7-oxononanoate aminotransferase;
KM seborrhoea; dermatitis; ds.
XX Escherichia coli DSM498.
XX
FH Key
FT promoter
FT
FT Location/Qualifiers
FT   1..96
FT   /tag= a
FT   /function= "promoter plac"
FT   /evidence= EXPERIMENTAL
FT   23..28
FT   /tag= b
FT   /standard_name= "promoter plac"
FT   45..50
FT   /tag= c
FT   /evidence= EXPERIMENTAL
FT   /standard_name= "promoter plac"
FT   105..109
FT   /tag= d
FT   /evidence= EXPERIMENTAL
FT   /standard_name= "bioB RBS no. 9"
FT   117..1157
FT   /tag= e
FT   /product= "biotin synthase"
FT   /evidence= EXPERIMENTAL
FT   /gene= "bioB"
FT   /number= 1
FT   1141..1146
FT   /tag= f
FT   /standard_name= "bioF RBS"
FT   1154..2311
FT   /tag= g
FT   /EC_number= 2.3.1.47
FT   /product= "KAPA synthase"
FT   /evidence= EXPERIMENTAL
FT   /gene= "bioF"
FT   /number= 2
FT   /standard_name= "8-amino-7-oxononanoate synthase"
FT   2284..2288
FT   /tag= h
FT   /standard_name= "bioC RBS"
FT   2295..3050
FT   /tag= i
FT   /function= "involved in pimeloyl-CoA synthesis"
FT   /product= "protein"
FT   /gene= "bioC"
FT   /number= 3
FT   3030..3033
FT   /tag= j
FT   /standard_name= "bioD RBS"
FT   3043..3753
FT   /tag= k
FT   /EC_number= 6.3.3.3
FT   /product= "DTB synthase"
FT   /evidence= EXPERIMENTAL
FT   /gene= "bioD15"
FT   /number= 4
FT   /standard_name= "dethiobiotin synthase"
FT   3712..3750
FT   /tag= l
FT   /note= "bioD15 substitution"
FT   3742..3746
FT   /tag= m
FT   /standard_name= "bioA RBS"
FT   3750..5039
FT   /tag= n
FT   /EC_number= 2.6.1.62
FT   /product= "DAPA synthase"
FT   /evidence= EXPERIMENTAL
FT   /gene= "bioA"
FT   /number= 5
FT   /standard_name= "S-adenosyl-L-methionine: 8-amino-

```

```

FT   RBS
FT   5088..5093
FT   /tag= o
FT   /standard_name= "ORF1 RBS"
FT   5098..5574
FT   CDS
FT   /tag= p
FT   /function= "unknown, involved in biotin synthesis"
FT   /product= "protein"
FT   /evidence= EXPERIMENTAL
FT   /gene= "ORF1"
FT   /number= 6
FT   5583..5644
FT   /tag= q
FT   /standard_name= "rho-independent transcriptional
FT   terminator"
FT   5583..5605
FT   /tag= r
FT   stem_loop
FT   5583..5605
FT   /tag= r
FT   WO9408023-A.
FT   14-APR-1994.
FT   PD
FT   14-APR-1994.
FT   PD
FT   01-OCT-1993; 93WO-EP02688.
FT   PF
FT   02-OCT-1992; 92CH-0003124.
FT   PR
FT   15-JUL-1993; 93CH-0002134.
FT   XX
FT   (LONZ ) LONZA AG.
FT   PA
FT   XX
FT   Birch O, Brass J, Fuhrmann M, Shaw N;
FT   PI
FT   XX
FT   WPI: 1994-135587/16.
FT   DR
FT   P-PSDB; AAR51883, AAR51884, AAR51885, AAR51886, AAR51887, AAR63121.
FT   XX
FT   Biotechnological biotin prodn. using enterobacterial biotin-gene
FT   PT
FT   - providing vitamin H in high yield
FT   XX
FT   Claim 1: Fig 6, Page 47-55 and 60-65; 92pp; German.
FT   PS
FT   The sequence is derived from plasmid pB030A-15/9 contg. the
FT   CC
FT   bioB, bioF, bioC, bioD and bioA genes responsible for biosynthesis
FT   CC
FT   of biotin, arranged in a transcription unit. Microorganisms
FT   CC
FT   contg. these DNA fragments or plasmids may be used in the prodn.
FT   CC
FT   of biotin. Biotin (Vitamin H) may prevent seborrhoea, dermatitis,
FT   CC
FT   loss of appetite and tiredness.
FT   XX
FT   SO
FT   Sequence 5872 BP; 1318 A; 1552 C; 1695 G; 1307 T; 0 other;
FT
FT   Query Match
FT   Best local Similarity 100.0%; Score 25; DB 15; Length 5872;
FT   Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
FT
FT   QY
FT   1 ggaatcgctgttgaagcgacgacg 25
FT   |||||||||||||||||||||||||
FT   Db 179 ggaatcgctgttgaagcgacgacg 203
FT
FT   RESULT 5
FT   AAX20559/C
FT   ID AAX20559 standard; DNA: 5504 BP.
FT   AC
FT   AAX20559;
FT   XX
FT   05-MAY-1999 (first entry)
FT   XX
FT   Polynucleotide sequence from the genome of Treponema pallidum.
FT   DE
FT   Treponema pallidum infection; syphilis; Borrelia infection; animal;
FT   KW
FT   enzyme production; ds.
FT   XX
FT   OS
FT   Treponema pallidum.
FT   XX

```

PR 08-JUL-1999; 99DE-1031435.
PR 08-JUL-1999; 99DE-1031443.
PR 08-JUL-1999; 99DE-1031453.
PR 08-JUL-1999; 99DE-1031457.
PR 08-JUL-1999; 99DE-1031465.
PR 08-JUL-1999; 99DE-1031478.
PR 08-JUL-1999; 99DE-1031510.
PR 08-JUL-1999; 99DE-1031541.
PR 08-JUL-1999; 99DE-1031573.
PR 08-JUL-1999; 99DE-1031592.
PR 08-JUL-1999; 99DE-1031632.
PR 08-JUL-1999; 99DE-1031634.
PR 08-JUL-1999; 99DE-1031636.
PR 09-JUL-1999; 99DE-1032125.
PR 09-JUL-1999; 99DE-1032126.
PR 09-JUL-1999; 99DE-1032130.
PR 09-JUL-1999; 99DE-1032186.
PR 09-JUL-1999; 99DE-1032206.
PR 09-JUL-1999; 99DE-1032227.
PR 09-JUL-1999; 99DE-1032228.
PR 09-JUL-1999; 99DE-1032229.
PR 14-JUL-1999; 99DE-1032920.
PR 14-JUL-1999; 99DE-1032922.
PR 14-JUL-1999; 99DE-1032926.
PR 14-JUL-1999; 99DE-1032928.
PR 14-JUL-1999; 99DE-1033004.
PR 14-JUL-1999; 99DE-1033005.
PR 12-AUG-1999; 99DS-0148613.
PR 27-AUG-1999; 99DE-1040764.
PR 27-AUG-1999; 99DE-1040765.
PR 27-AUG-1999; 99DE-1040766.
PR 27-AUG-1999; 99DE-1040767.
PR 27-AUG-1999; 99DE-1040832.
PR 31-AUG-1999; 99DE-1041378.
PR 31-AUG-1999; 99DE-1041379.
PR 31-AUG-1999; 99DE-1041380.
PR 31-AUG-1999; 99DE-1041394.
PR 31-AUG-1999; 99DE-1041396.
PR 03-SEP-1999; 99DE-1042076.
PR 03-SEP-1999; 99DE-1042077.
PR 03-SEP-1999; 99DE-1042079.
PR 03-SEP-1999; 99DE-1042086.
PR 03-SEP-1999; 99DE-1042087.
PR 03-SEP-1999; 99DE-1042088.
PR 03-SEP-1999; 99DE-1042095.
PR 03-SEP-1999; 99DE-1042124.
PR 03-SEP-1999; 99DE-1042129.
PR 09-MAR-2000; 2000US-0187970.

XX (BADI) BASF AG.
XX
XX
XX Pompejus M., Kreager B., Schroeder H., zelder O., Habernauer G;
PI WPl: 2001-137957/14.
DR P-PSDB: AAB79808.
XX
XX Nucleic acids from Corynebacterium glutamicum encoding metabolic
PT pathway proteins, useful for producing fine chemicals in
PT microorganisms, including organic acids, nonproteinogenic amino acids,
PT and purine and pyrimidine bases -
XX
XX
XX Claim 3; Page 676-677; 1737pp; English.
XX
CC AAT71753 to AAT72330 encode the Corynebacterium glutamicum metabolic
CC pathway (MP) proteins given in AAB79634 to AAB80211. The C. glutamicum
CC MP nucleic acids are useful for the production of fine chemicals
CC in microorganisms, including organic acids, nonproteinogenic amino
CC acids, purine and pyrimidine bases, nucleosides, nucleotides, lipids,
CC saturated and unsaturated fatty acids, diols, carbohydrates, aromatic
CC compounds, vitamins, cofactors, polyketides and enzymes.
XX
XX Sequence 924BP; 214 A; 252 C; 239 G; 219 T; 0 other;

Query Match 64.0%; Score 16; DB 22; Length 924;
Best Local Similarity 100.0%; Pred. No. 3.6;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 8 cgtgtgaagcgcgc 23
|||||
Db 884 CTGTTGAGCGCAGC 869

RESULT 7
AAZ23024/c
ID AAZ23024 standard; cDNA; 1689 BP.
XX
AC AAZ23024;
XX
DT 17-JAN-2000 (first entry)
XX
DE Rat kd312 polypeptide encoding cDNA.
XX
KW kd312 polypeptide; cancer; immunodeficiency disorder; AIDS; stroke;
KW heart attack; head trauma; neurodegenerative disease; rat;
KW Parkinson's disease; Alzheimer's disease; ss.
XX
OS Rattus sp.
XX
PN MO9950288-A2.
XX
PD 07-OCT-1999.
XX
PF 30-MAR-1999; 99WO-US06993.
XX
PR 31-MAR-1998; 98US-0053374.
XX
PA (AMGE-) AMGEN INC.
XX
PI Yen K;
XX
DR WPI; 1999-601322/51.
DR P-PSDB; AA42694.
XX
XX
PT kd312 polypeptides useful for treating diseases and disorders
PT associated with alterations in cell proliferation and cell death
XX
PS Claim 2; Fig 9; 85pp; English.
XX
CC The invention provides nucleic acid molecules encoding human and rat
CC kd312 polypeptides. The kd312 polypeptides can be expressed by standard
CC recombinant methodology. The kd312 sequences, and the antibodies against
CC the proteins may be used to treat or diagnose the presence or progression
CC of various diseases, e.g. cancer, immunodeficiency disorders (e.g. AIDS),
CC stroke, heart attack, head trauma, and neurodegenerative diseases (e.g.
CC Parkinson's disease and Alzheimer's disease). The present sequence
CC represents the rat kd312 cDNA sequence.
XX
SQ Sequence 1689 BP; 417 A; 462 C; 434 G; 376 T; 0 other;

Query Match 60.0%; Score 15; DB 20; Length 1689;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 ggaatcgtcgtttga 15
|||||
Db 495 GCATCTGCTTTGA 481

RESULT 8
AAC44132/c
ID AAC44132 standard; DNA; 2022 BP.
XX
AC AAC44132;
XX

DT 18-OCT-2000 (first entry)
XX
DE Zea mays DNA fragment SEQ ID NO: 41750.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway; metabolic;
KW pathway; promoter; termination sequence; corn; ss.
XX
OS Zea mays subsp. mays.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.

```

PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141844.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142972.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 27-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145919.
PR 02-AUG-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 09-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 10-AUG-1999; 99US-0147935.
PR 11-AUG-1999; 99US-0148171.
PR 12-AUG-1999; 99US-0148319.
PR 13-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0148684.
PR 17-AUG-1999; 99US-0149368.
PR 18-AUG-1999; 99US-0149375.
PR 20-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 23-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 25-AUG-1999; 99US-0149930.
PR 26-AUG-1999; 99US-0150566.
PR 27-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.

```

```

PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151308.
PR 07-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154379.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155386.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 22-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 25-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

Query Match 60.0%; Score 15; DB 21; Length 2022;

Best Local Similarity 100.0%; Pred. No. 13;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 tctgctgttgagc 18

Db 1268 TCTGCTGTTGAAC 1254

RESULT 9

ID AA23025/c

AA23025 standard; DNA; 3079 BP.

AC AA23025;

DT 17-JAN-2000 (first entry)

DE Rat Kd312 genomic DNA sequence.

XX Kd312 polypeptide; cancer; immunodeficiency disorder; ALDs; stroke;

KW heart attack; head trauma; neurodegenerative disease; rat;

KW Parkinson's disease; Alzheimer's disease; ss.

XX Rattus sp.
 OS WO950288-A2.
 XX 07-OCT-1999.
 PD
 XX 30-MAR-1999; 99WO-US06993.
 PF
 XX 31-MAR-1998; 98US-0053374.
 PR
 XX (AMGE-) AMGEN INC.
 PA
 XX Yen K.
 PI
 XX WPI: 1999-601322/51.
 DR
 XX P-PSDB: AAY42694.
 DR
 XX
 PT kd312 polypeptides useful for treating diseases and disorders
 associated with alterations in cell proliferation and cell death
 XX
 PS Claim 2; Fig 7; 85pp; English.
 CC
 CC The invention provides nucleic acid molecules encoding human and rat
 CC kd312 polypeptides. The kd312 polypeptides can be expressed by standard
 CC recombinant methodology. The kd312 sequences, and the antibodies against
 CC the proteins may be used to treat or diagnose the presence or progression
 CC of various diseases, e.g. cancer, immunodeficiency disorders (e.g. AIDS),
 CC stroke, heart attack, head trauma, and neurodegenerative diseases (e.g.
 CC Parkinson's disease and Alzheimer's disease). The present sequence
 CC represents the rat kd312 genomic DNA sequence.
 CC
 SO Sequence 3079 BP; 708 A; 843 C; 829 G; 699 T; 0 other;

Query Match 60.0%; Score 15; DB 20; Length 3079;
 Best Local Similarity 100.0%; Pred. NO. 13;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ggatcgtctgttga 15
 |||||||
 DB 803 GGATCTGCTGTTGA 789

RESULT 10
 AAF65487/C
 ID AAF65487 standard; CDNA; 336 BP.
 XX
 AC AAF65487;
 XX
 DT 09-APR-2001 (first entry)
 XX
 DE Novel human polynucleotide, SEQ ID NO: 1243.
 XX
 XX Human: cytostatic; gene therapy: colon cancer; prostate cancer;
 KW breast cancer; lung cancer; cancer detection; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200102568-A2.
 XX
 PD 11-JAN-2001.
 XX
 PF 30-JUN-2000; 2000WO-US18374.
 XX
 PR 02-JUL-1999; 99US-0142310.
 PR 02-JUL-1999; 99US-0142311.
 XX
 PA (CHIR) CHIRON CORP.
 PA (HYSE-) HYSEQ INC.
 PI Williams LT, Escobedo J, Innis MA, Garcia PD, Klinger J, Kassam A,
 PI Reinhard C, Randazzo F, Kennedy GC, Pot D, Lamson G, Drmanac R;

PI Orkenjakov R, Drmanac S, Dickson M, Labat I, Leshkowitz D;
 PI Kita D, Garcia V, Jones LM, Strache-Grain B;
 XX WPI: 2001-091805/10.
 DR
 XX
 PT library of polynucleotides for diagnosing a cancerous state of a
 PT mammalian cell and detecting cancer, particularly of the colon or
 PT prostate, comprises 3351 human polynucleotide sequences -
 PS
 XX Claim 9; Page 722; 1046pp; English.

CC The present sequence is one of 3351 sequences in a library of human
 CC polynucleotides. The library is used to detect differentially expressed
 CC genes correlated with a cancerous state of a mammalian cell and can
 CC detect colon, prostate, breast and lung cancer. The library can be used
 CC to produce probes for detection of mRNA and to produce additional copies
 CC of the polynucleotides. The probes can be used for chromosome mapping of
 CC the polynucleotide and for detection of transcription levels. Ribozymes
 CC or antisense oligonucleotides can be generated. The polynucleotides and
 CC their gene products are used as genetic or biochemical markers (e.g. in
 CC blood or tissues) that will detect the earliest changes along the
 CC carcinogenesis pathway and/or monitor the efficacy of therapies and
 CC preventive interventions. The polynucleotides, polypeptides and
 CC antibodies against them can be used in pharmaceutical compositions to
 CC treat the cancers and proliferative disorders such as neoplasia,
 CC dysplasia and hyperplasia.

Sequence 336 BP; 97 A; 84 C; 59 G; 96 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 336;
 Best Local Similarity 100.0%; Pred. No. 46;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 atctgctgttga 16
 |||||||
 DB 159 ATCTGCTGTTGAA 146

RESULT 11
 AAV87319/C
 ID AAV87319 standard; CDNA; 379 BP.
 XX
 AC AAV87319;
 XX
 DT 27-APR-1999 (first entry)
 XX
 DE EST clone BP797.
 XX
 KW Expressed sequence tag; secreted protein; haematopoiesis regulator;
 KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
 KW chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;
 KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.
 XX
 XX Homo sapiens.
 OS
 XX
 PN WO9845435-A2.
 XX
 PD 15-OCT-1998.
 XX
 PF 10-APR-1998; 98WO-US06954.
 XX
 PR 10-APR-1997; 97US-0835913.
 XX
 PA (GENY) GENETICS INST INC.
 XX
 PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
 PI Racine LA, Spaulding V, Treacy M;
 XX WPI: 1999-070076/06.
 DR
 XX
 PT New polynucleotides encoding human secreted proteins - derived from
 PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,

PT ovary, pituitary, retina and colon cDNA libraries
 PS Claim 1; Page 527; 633pp; English.

CC This sequence represents an expressed sequence tag (EST), and is a
 CC polynucleotide of the invention. The polynucleotides of the invention are
 CC all secreted EST sequences isolated from a variety of human tissue
 CC sources. The EST sequences and proteins encoded by them are predicted to
 CC have useful biological activities which would make them suitable for
 CC treating, preventing or ameliorating medical conditions in humans and
 CC animals, although no supporting data is given. Suggested activities
 CC include nutritional activity, immune stimulating or suppressing activity,
 CC haemopoiesis regulating activity, tissue growth activity,
 CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC activity, thrombolytic activity, receptor/ligand activity, anti-inflammatory
 CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition
 CC therapy. The EST sequences are also stated to be useful for gene
 CC therapy.

CC Sequence 379 BP; 124 A; 68 C; 70 G; 116 T; 1 other;

Query Match Best Local Similarity 56.0%; Score 14; DB 20; Length 379;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 gatcgtcgtttga 15
 Db 129 GATCTGCTTTTGA 116

RESULT 12
 ID AAA44124/c
 AC AAA44124 standard; cDNA; 619 BP.

XX AAA44124;

DT 21-AUG-2000 (first entry)

DE Human secreted expressed sequence tag SEQ ID NO:699.

XX Human: mouse; chicken; rat; secreted expressed sequence tag; SEST;
 KW expressed sequence tag; EST; probe; chemotactic; proliferative;
 KW immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;
 KW antitumor; antiinflammatory; cytostatic; antibacterial; antifungal;
 KW antiviral; antidiabetic; antiasthmatic; vulnery; antiparkinsonian;
 KW antitumor; osteopathic; neuroprotective; nocitropic; antipsoriatic;
 KW cerebroprotective; anticonvulsant; antidepressant; gene therapy;
 KW vaccine; autoimmune disorder; multiple sclerosis; allergic condition;
 KW insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;
 KW lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;
 KW central nervous system disorder; Alzheimer's disease; stroke;
 KW Parkinson's disease; Huntington's disease; coagulation disorder;
 KW haemophilia; thrombosis; inflammatory disorder; Crohn's disease;
 KW tumour; infection; depression; psoriasis; ss.

OS Homo sapiens.

PN WO200021991-A1.

PD 20-APR-2000.

PF 15-OCT-1999; 99WO-US24206.

PR 15-OCT-1998; 98US-014436.

PA (GEM) GENETICS INST INC.

PI Jacobs K, McCoy JM, LaValle ER, Collins-Racie LA, Evans C;

XX Metberg D, Treacy M, Bowman MR;

XX WPI; 2000-317938/27.

PT Isolated polynucleotides, and encoded proteins, comprising secreted
 PT expressed sequence tags (SESTs), useful for treating various disorders
 PT such as autoimmune, infectious, and central nervous system disorders -
 PS Claim 1; Page 383; 803pp; English.

CC AAA43426 to AAA45925 represent specifically claimed secreted expressed
 CC sequence tags (SESTs), isolated from human, mouse, chicken and rat
 CC tissue sources. The SESTs can have a range of activities depending on
 CC the tissues they were isolated from. The activities include:
 CC chemotactic; proliferative; immunomodulatory; haematopoietic;
 CC cyostatic; analgesic; haemostatic; thrombolytic; antiinflammatory;
 CC antiasthmatic; antibacterial; antifungal; antiviral; antidiabetic;
 CC neurotropic; antiparkinsonian; antipsoriatic; cerebroprotective;
 CC anticonvulsant; and antidepressant. The SESTs can be used for gene
 CC therapy and in vaccines. The SESTs are useful as probes for the
 CC identification and isolation of full-length cDNAs and genomic DNA
 CC molecules which correspond to the SESTs. Proteins encoded by the SESTs
 CC are useful in assays for determining biological activity and raising
 CC antibodies. They may be useful for treatment of autoimmune disorders
 CC (multiple sclerosis, insulin dependent diabetes), allergic conditions
 CC (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,
 CC osteoporosis, osteoarthritis, central nervous system disorders
 CC (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation
 CC disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's
 CC disease), tumours, bacterial, fungal or viral infections, depression and
 CC psoriasis. AAA45926 to AAA45931 represent linker variants which are given
 CC in the exemplification of the present invention.

CC Sequence 619 BP; 193 A; 123 C; 125 G; 178 T; 0 other;

Query Match Best Local Similarity 56.0%; Score 14; DB 21; Length 619;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 atctgctgtttga 16
 Db 43 ATCTGCTTTTGA 30

RESULT 13

ID AA200347/c

AC AA200347 standard; DNA; 753 BP.

XX AA200347;

DT 26-OCT-1999 (first entry)

DE Nucleotide sequence of human hSFATP3.

KW Fatty acid transport protein; FATP; long chain fatty acid; LCFA; human;
 KW fatty acid; FATP biosynthesis; obesity; diabetes; heart disease; ss.

OS Homo sapiens.

PN WO9936537-A2.

PD 22-JUL-1999.

PF 14-JAN-1999; 99WO-US00182.

PR 14-JAN-1999; 99US-0232201.

PA 15-JAN-1998; 98US-0071374.

PR 20-JUL-1998; 98US-0093491.

PR 04-DEC-1998; 98US-0110941.

PR 14-JAN-1999; 99US-0232195.

PR 14-JAN-1999; 99US-0232197.

XX 14-JAN-1999; 99US-0232200.

PA (MILL-) MILLENNIUM PHARM INC.

XX (WHED) WHITEHEAD INST BIOMEDICAL RES.

PI Gimeno RE, Hirsch DJ, Lodish HF, Stahl A, Tartaglia LA;
XX
XX WPI: 1999-444398/37.
DR P-PSDB; AAY14937.
XX
PT Fatty acid transport proteins and related polynucleotides, useful
PT for treating obesity, diabetes and heart disease
XX
XX
PS Example 1; Fig 16; 255pp; English.
XX
XX
CC The invention provides a family of fatty acid transport proteins (FATPs)
CC that mediate transport of long chain fatty acids (LCFAs) across cell
CC membranes into cells. Human and murine FATP proteins and nucleic acids
CC encoding the proteins are provided. The FATP proteins can be produced
CC by standard recombinant methodology. Fatty acid uptake by cells can be
CC modulated by modulating biosynthesis of FATP proteins especially FATP6.
CC In particular, antisense oligonucleotides can be used to modulate FATP
CC biosynthesis. Modulation of FATP6 is useful for inhibiting fatty acid
CC uptake in cardiac muscle of humans. Agents can be directed to cardiac
CC muscle or liver by administration of a complex of the agent and a FATP6
CC binding moiety. DNA encoding FATP proteins can be used as a reference
CC used in detecting variant alleles or homologues. Altering the LCFA uptake
CC by administering an inhibitor or enhancer of FATP transport function in
CC the small intestine can decrease or increase calories available as fats,
CC and can decrease or increase circulating fatty acids. Blocking the
CC function of FATP4 and also FATP2, is useful for treating obesity,
CC diabetes and heart disease.
XX
XX Sequence 753 BP; 175 A; 206 C; 206 G; 166 T; 0 other;

Query Match	56.0%	Score 14:	DB 20;	Length 753;
Best Local Similarity	100.0%	Pred. No. 46;		
Matches 14;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	4	tctgctgttgaa	17
Db	436	TCTGCTGTTGAAG	423

RESULT	14
AAZ00358/c	
ID	AAZ00358 standard; DNA; 1173 BP

AC	AAZ00358;
XX	
DT	26-OCT-1999 (first entry)

XX Fatty acid transport protein; FATP; long chain fatty acid; LCFA; human;
KW fatty acid; FATP biosynthesis; obesity; diabetes; heart disease; ss.
KW

OS	Homo sapiens
XX	
PN	W09936537-A2

XX	14 - JAN - 1999;	99WO-US00182
PF		
XX		

PR	20-JUL-1998;	98US-0093491
PR	04-DEC-1998;	98US-0110941
PR	14-JAN-1999;	99US-0232195

XX (MILL-) MILLENNIUM PHARM INC.
PA (WHED) WHITEHEAD INST BIOMEDICAL RES
PA

PI Giмено RE, Hirsch DJ, Lodish HF, Stahl A, Tartaglia LA;
 DR WPI: 1999-444398/37.
 DR P-PSDB; AAY14948.
 XX
 PT Fatty acid transport proteins and related polynucleotides, useful
 PT for treating obesity, diabetes and heart disease
 XX
 PS Claim 31; Fig 48; 255pp; English.
 XX
 CC The invention provides a family of fatty acid transport proteins (FATPs)
 CC that mediate transport of long chain fatty acids (LCFAs) across cell
 CC membranes into cells. Human and murine FATP proteins and nucleic acids
 CC encoding the proteins are provided. The FATP proteins can be produced
 CC by standard recombinant methodology. Fatty acid uptake by cells can be
 CC modulated by modulating biosynthesis of FATP proteins especially FATP6.
 CC In particular, antisense oligonucleotides can be used to modulate FATP
 CC biosynthesis. Modulation of FATP6 is useful for inhibiting fatty acid
 CC uptake in cardiac muscle of humans. Agents can be directed to cardiac
 CC muscle or liver by administration of a complex of the agent and a FATP6
 CC binding moiety. DNA encoding FATP proteins can be used as a reference
 CC used in detecting variant alleles or homologues. Altering the LCFa uptake
 CC by administering an inhibitor or enhancer of FATP transport function in
 CC the small intestine can decrease or increase calories available as fats,
 CC and can decrease or increase circulating fatty acids. Blocking the
 CC function of FATP4 and also FATP2, is useful for treating obesity,
 CC diabetes and heart disease.
 XX
 SO Sequence 1173 BP; 247 A; 327 C; 351 G; 248 T; 0 other;

Query Match	56.0%	Score 14	DB 20	Length 1173
Best Local Similarity	100.0%	Pred. No. 47		
Matches 14	Conservative 0	Mismatches 0	Indels 0	Gaps 0

QY	4	tctgctgttgaag	17
Db	853	TCTGCTGTTGAAG	840

RESULT	15
AAA50896/c	
ID	AAA50896 standard; cDNA; 1244 BP.

AC AAA50896;
XX
DT 28-SEP-2000 (first entry)

XX Enzyme; phosphatidylinositol-4-phosphate kinase; PI;
KW phosphatidylinositol metabolism; signal transduction; membrane a

XX Glycine max
OS
XX

FT	/*tag= a
FT	/partial
FT	/product= "ssm.pk0068.c5"

XX 22-JUN-2000.
PD
XX

PA (DUPO) DU PONT DE NEMOURS & CO E I
XX
PR 18-DEC-1998; 98US-0112925.
ZAZ

XX (DUPO) DU PONT DE NEMOURS & CO E I
PA
XX

PI Allen SM, Kinney AJ, Miao G, Rafalski JA, Sakai H, Weng Z;
 XX WPI: 2000-431597/37.
 DR P-PSDB: AAY9751.
 XX
 XX New isolated polynucleotide encoding phosphatidylinositol metabolism
 PT protein is useful for producing transgenic plants with an altered level
 PT of phosphatidylinositol -
 XX
 XX Claim 2: Page 50-51; 65pp; English.
 XX
 XX Phosphatidylinositol, PI, may be modified in cells to have roles in
 CC signal transduction, membrane associated protein anchoring and regulation
 CC of membrane trafficking pathways. The enzyme phosphatidylinositol 4-
 CC phosphate 5-kinase, converts phosphatidylinositol-4-phosphate to
 CC phosphatidylinositol-4,5-diphosphate. Phosphatidylinositol-4,5-
 CC diphosphate is thought to play a role in vacuole homeostasis regulation.
 CC The present sequence is the partial coding sequence of the soybean PI
 CC metabolism protein clone, ssm.pk0068.c5, isolated from an expressed
 CC sequence tag, EST. The present sequence may be used to create transgenic
 CC plants which express phosphatidylinositol-4-phosphate kinase, and hence
 CC be useful for modifying plant phosphatidylinositols.
 CC
 SQ Sequence 1244 BP; 392 A; 207 C; 274 G; 371 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 1244;
 Best Local Similarity 100.0%; Pred. No. 47;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 gttgaagcagcagc 23
 |||||
 Db 886 GTTGAAGCGCAGC 873

RESULT 16
 AA004684
 ID AA004684 standard; DNA: 1536 BP.
 XX
 AC AA004684;
 XX
 DT 05-OCT-1990 (first entry)
 XX
 DE Salt/Kpni fragment of pUOI contg. haloacetate dehalogenase H-1.
 XX
 KW Haloacetate dehalogenase; H-1; H-2; ss.
 XX
 OS
 XX Moraxell sp. B.
 XX
 FH Key Location/Qualifiers
 FT 413..1296
 FT /*tag= a
 FT CDS /label=H-1_gene
 XX
 XX JP02109981-A.
 XX
 PD 23-APR-1990.
 XX
 PF 08-APR-1988; 88JP-0085393.
 XX
 PR 08-APR-1988; 88JP-0085393.
 XX
 PA (MTTN) MITSUBISHI GAS CHEM KK.
 DR WPI: 1990-168356/22.
 DR P-PSDB: AAR06435.
 XX
 PT Haloacetate dehalogenase coding gene - used to raise
 PT decomposition activity of aragane halogenated cpds.
 XX
 PS Claim 1; Fig 2; 8pp; Japanese.
 XX
 CC The DNA can be ligated into a vector which can then be used to

CC transform E.coli for the prodn. of H-1. The protein is used in
 CC the decomposition of organic halogenated cpds.
 CC See also AA004685.
 CC
 XX

SO Sequence 1536 BP; 315 A; 445 C; 449 G; 326 T; 1 other;

Query Match 56.0%; Score 14; DB 11; Length 1536;
 Best Local Similarity 100.0%; Pred. No. 47;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 ttgaagcagcagcag 25
 |||||
 Db 320 ttgaagcagcagcag 333

RESULT 17
 AA200367/c
 ID AA200367 standard; DNA: 1998 BP.
 XX
 AC AA200367;
 XX
 DT 26-OCT-1999 (first entry)
 XX
 DE Nucleotide sequence of murine mmFATP3.
 XX
 KW Fatty acid transport protein; FATP; long chain fatty acid; LCFA; murine;
 KW fatty acid; FATP biosynthesis; obesity; diabetes; heart disease; ss.
 XX
 OS Mus sp.
 XX
 PN WO936537-A2.
 PD 22-JUL-1999.
 XX
 PF 14-JAN-1999; 99WO-US00182.
 XX
 PR 14-JAN-1999; 99US-0232201.
 PR 15-JAN-1998; 98US-0071374.
 PR 20-JUL-1998; 98US-0093491.
 PR 04-DEC-1998; 98US-0110941.
 PR 14-JAN-1999; 99US-0232195.
 PR 14-JAN-1999; 99US-0232197.
 PR 14-JAN-1999; 99US-0232200.
 XX

PA (MILL-) MILLENNIUM PHARM INC.
 PA (WHED) WHITEHEAD INST BIOMEDICAL RES.
 XX
 PI Gimeno RE, Hirsch DJ, Lodish HF, Stahl A, Tartaglia LA;
 XX

WPI: 1999-444398/37.
 DR P-PSDB: AAY14957.
 XX

PT Fatty acid transport proteins and related polynucleotides, useful
 PT for treating obesity, diabetes and heart disease
 XX

Example 1; Fig 66; 255pp; English.

CC The invention provides a family of fatty acid transport proteins (FATPs)
 CC that mediate transport of long chain fatty acids (LCFAs) across cell
 CC membranes into cells. Human and murine FATP proteins and nucleic acids
 CC encoding the proteins are provided. The FATP proteins can be produced
 CC by standard recombinant methodology. Fatty acid uptake by cells can be
 CC modulated by modulating biosynthesis of FATP proteins especially FATP6.
 CC In particular, antisense oligonucleotides of FATP proteins can be used to modulate FATP
 CC biosynthesis. Modulation of FATP6 is useful for inhibiting fatty acid
 CC uptake in cardiac muscle of humans. Agents can be directed to cardiac
 CC muscle or liver by administration of a complex of the agent and a FATP6
 CC binding moiety. DNA encoding FATP proteins can be used as a reference
 CC used in detecting variant alleles or homologues. Altering the LCFA uptake
 CC by administering an inhibitor or enhancer of FATP transport function in
 CC the small intestine can decrease or increase calories available as fats,
 CC and can decrease or increase circulating fatty acids. Blocking the

CC function of FATP4 and also FATP2, is useful for treating obesity,
 CC diabetes and heart disease.
 XX
 XX Sequence 1998 BP; 422 A; 567 C; 597 G; 412 T; 0 other;

Query Match 56.0%; Score 14; DB 20; Length 1998;
 Best Local Similarity 100.0%; Pred. No. 47;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 tctgctgttgaag 17
 |||||
 Db 1690 TCTGCTGTTGAAG 1677

RESULT 18
 AA200343/C
 ID AA200343 standard; DNA: 2087 BP.

XX AA200343;
 XX 26-OCT-1999 (first entry)
 DE Nucleotide sequence of murine mmFATP3.
 XX Fatty acid transport protein; FATP; long chain fatty acid; LCFA; murine;
 KW fatty acid; FATP biosynthesis; obesity; diabetes; heart disease; ss.
 XX
 OS Mus musculus.

PN W09936537-A2.

PD 22-JUL-1999.

PF 14-JAN-1999; 99WO-US00182.

PR 14-JAN-1999; 99US-0232201.

PR 15-JAN-1998; 98US-0071374.

PR 20-JUL-1998; 98US-0093491.

PR 04-DEC-1998; 98US-0110941.

PR 14-JAN-1999; 99US-0232195.

PR 14-JAN-1999; 99US-0232197.

PR 14-JAN-1999; 99US-0232200.

PA (MILL-) MILLENNIUM PHARM INC.

PA (WHED) WHITEHEAD INST BIOMEDICAL RES.

PI Glimeno RE, Hirsch DJ, Lodish HF, Stahl A, Tartaglia LA;

PI WPI: 1999-444398/37.

DR P-PSDB; AAY14933.

DR Fatty acid transport proteins and related polynucleotides, useful

PT for treating obesity, diabetes and heart disease

PS Example 1; Fig 8; 255pp; English.

XX The invention provides a family of fatty acid transport proteins (FATPs)
 CC that mediate transport of long chain fatty acids (LCFAs) across cell
 CC membranes into cells. Human and murine FATP proteins and nucleic acids
 CC encoding the proteins are provided. The FATP proteins can be produced
 CC by standard recombinant methodology. Fatty acid uptake by cells can be
 CC modulated by modulating biosynthesis of FATP proteins especially FATP6.
 CC In particular, antisense oligonucleotides can be used to modulate FATP
 CC biosynthesis. Modulation of FATP6 is useful for inhibiting fatty acid
 CC uptake in cardiac muscle of humans. Agents can be directed to cardiac
 CC muscle or liver by administration of a complex of the agent and a FATP6
 CC binding moiety. DNA encoding FATP proteins can be used as a reference
 CC used in detecting variant alleles or homologues. Altering the LCFA uptake
 CC by administering an inhibitor or enhancer of FATP transport function in
 CC the small intestine can decrease or increase calories available as fats,
 CC and can decrease or increase circulating fatty acids. Blocking the
 CC function of FATP4 and also FATP2, is useful for treating obesity,

CC diabetes and heart disease.
 XX
 XX Sequence 2087 BP; 441 A; 596 C; 624 G; 426 T; 0 other;

Query Match 56.0%; Score 14; DB 20; Length 2087;
 Best Local Similarity 100.0%; Pred. No. 47;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 tctgctgttgaag 17
 |||||
 Db 1779 TCTGCTGTTGAAG 1766

RESULT 19
 AA200379/C
 ID AA200379 standard; DNA: 2166 BP.

XX AA200379;
 XX 26-OCT-1999 (first entry)
 DE Nucleotide sequence of human hsFATP5 gene.
 XX Fatty acid transport protein; FATP; long chain fatty acid; LCFA; human;
 KW fatty acid; FATP biosynthesis; obesity; diabetes; heart disease; ss.
 XX
 OS Homo sapiens.

PN W09936537-A2.

PD 22-JUL-1999.

PF 14-JAN-1999; 99WO-US00182.

PR 14-JAN-1999; 99US-0232201.

PR 15-JAN-1998; 98US-0071374.

PR 20-JUL-1998; 98US-0093491.

PR 04-DEC-1998; 98US-0110941.

PR 14-JAN-1999; 99US-0232195.

PR 14-JAN-1999; 99US-0232197.

PR 14-JAN-1999; 99US-0232200.

PA (MILL-) MILLENNIUM PHARM INC.

PA (WHED) WHITEHEAD INST BIOMEDICAL RES.

PI Glimeno RE, Hirsch DJ, Lodish HF, Stahl A, Tartaglia LA;

PI WPI: 1999-444398/37.

DR P-PSDB; AAY14969.

DR Fatty acid transport proteins and related polynucleotides, useful

PT for treating obesity, diabetes and heart disease

PS Claim 31; Fig 94A-B; 255pp; English.

XX The invention provides a family of fatty acid transport proteins (FATPs)
 CC that mediate transport of long chain fatty acids (LCFAs) across cell
 CC membranes into cells. Human and murine FATP proteins and nucleic acids
 CC encoding the proteins are provided. The FATP proteins can be produced
 CC by standard recombinant methodology. Fatty acid uptake by cells can be
 CC modulated by modulating biosynthesis of FATP proteins especially FATP6.
 CC In particular, antisense oligonucleotides can be used to modulate FATP
 CC biosynthesis. Modulation of FATP6 is useful for inhibiting fatty acid
 CC uptake in cardiac muscle of humans. Agents can be directed to cardiac
 CC muscle or liver by administration of a complex of the agent and a FATP6
 CC binding moiety. DNA encoding FATP proteins can be used as a reference
 CC used in detecting variant alleles or homologues. Altering the LCFA uptake
 CC by administering an inhibitor or enhancer of FATP transport function in
 CC the small intestine can decrease or increase calories available as fats,
 CC and can decrease or increase circulating fatty acids. Blocking the
 CC function of FATP4 and also FATP2, is useful for treating obesity,
 CC diabetes and heart disease.

XX Sequence 2166 BP; 370 A; 682 C; 695 G; 419 T; 0 other;
 SQ
 Query Match
 Best Local Similarity 56.0%; Score 14; DB 20; Length 2166;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 4 tctgctgtttgaag 17
 |||
 Db 1987 TCTGCTGTTGAAG 1974
 RESULT 20
 AAF27417/C
 ID AAF27417 standard; cDNA: 2405 BP.
 XX
 AC AAF27417;
 XX
 DT 24-APR-2001 (first entry)
 XX
 DE Human fatty acid transporter PSEC67 cDNA.
 XX
 KW Human; fatty acid transporter; PSEC67; long-chain fatty acid uptake;
 KW oleic acid; drug screening; gene therapy; metabolic disorder;
 KW cardiomyopathy; skeletal muscle disorders; renal failure; ss.
 XX
 OS Homo sapiens.
 XX
 FN W020104301-A1.
 XX
 PD 18-JAN-2001.
 XX
 PE 07-JUL-2000; 2000MO-JP04549.
 XX
 PR 08-JUL-1999; 99JP-0194179.
 PR 18-OCT-1999; 99US-0159586.
 PR 25-APR-2000; 2000JP-0128993.
 XX
 PA (HELI-) HELIX RES INSTR.
 XX
 PI Morikawa N, Masuno Y, Ota T, Isogai T, Nishikawa T, Kawai Y;
 XX
 DR WPI; 2001-138349/14.
 DR P-PSDB; AAB60388.
 XX
 PT Fatty acid transporter protein and encoded gene PSEC67 cloned from
 PT human cDNA library, with activity of oleic acid incorporation, useful
 PT as target molecule of preventives or remedies of fatty-acid metabolic
 PT disorders -
 XX
 PS Claim 1: Page 44-48; 58pp; Japanese.
 XX
 CC The invention relates to a novel human fatty acid transporter,
 CC PSEC67 (AAB60388), and to cDNA encoding it (AAF27417). PSEC67 is
 CC responsible for the uptake of oleic acid into cells. The invention
 CC also relates to vectors and host cells comprising a PSEC67
 CC nucleic acid; the recombinant production of PSEC67; an antibody
 CC against PSEC67; methods of screening for compounds which can regulate
 CC the uptake of long-chain fatty acids into cells; and the compounds thus
 CC identified. The PSEC67 protein and the gene encoding it are useful as
 CC targets for the treatment or prevention of diseases associated with
 CC defective metabolism of long-chain fatty acids. Such diseases include
 CC cardiomyopathy, skeletal muscle disorders, or renal failure. PSEC67
 CC nucleic acids may be used in gene therapy for such disorders.
 CC The present sequence represents cDNA encoding PSEC67.
 CC
 XX
 SQ Sequence 2405 BP; 410 A; 735 C; 781 G; 479 T; 0 other;

QY 4 tctgctgtttgaag 17
 |||
 Db 2111 TCTGCTGTTGAAG 2098
 RESULT 21
 AA233977/C
 ID AA233977 standard; cDNA: 2574 BP.
 XX
 AC AA233977;
 XX
 DT 07-DEC-1999 (first entry)
 XX
 DE Human PRO703 nucleotide sequence.
 XX
 KW Human; PRO; EST; expressed sequence tag; PCR primer; hybridisation;
 KW probe; blood coagulation disorder; cancer; cellular adhesion disorder;
 KW secreted protein; transmembrane protein; ss.
 XX
 OS Homo sapiens.
 XX
 FN W09946281-A2.
 XX
 PD 16-SEP-1999.
 XX
 PE 08-MAR-1999; 99WO-US05028.
 XX
 PR 10-MAR-1998; 98US-0077450.
 PR 11-MAR-1998; 98US-0077632.
 PR 11-MAR-1998; 98US-0077641.
 PR 12-MAR-1998; 98US-0077649.
 PR 13-MAR-1998; 98US-0077791.
 PR 17-MAR-1998; 98US-0078004.
 PR 20-MAR-1998; 98US-0040220.
 PR 20-MAR-1998; 98US-0078886.
 PR 20-MAR-1998; 98US-0078910.
 PR 20-MAR-1998; 98US-0078936.
 PR 20-MAR-1998; 98US-0078939.
 PR 25-MAR-1998; 98US-0079294.
 PR 26-MAR-1998; 98US-0079656.
 PR 27-MAR-1998; 98US-0079663.
 PR 27-MAR-1998; 98US-0079664.
 PR 27-MAR-1998; 98US-0079689.
 PR 27-MAR-1998; 98US-0079728.
 PR 27-MAR-1998; 98US-0079786.
 PR 30-MAR-1998; 98US-0079920.
 PR 30-MAR-1998; 98US-0079923.
 PR 31-MAR-1998; 98US-0080105.
 PR 31-MAR-1998; 98US-0080107.
 PR 31-MAR-1998; 98US-0080165.
 PR 31-MAR-1998; 98US-0080194.
 PR 01-APR-1998; 98US-0080327.
 PR 01-APR-1998; 98US-0080328.
 PR 01-APR-1998; 98US-0080333.
 PR 01-APR-1998; 98US-0080334.
 PR 08-APR-1998; 98US-0081049.
 PR 08-APR-1998; 98US-0081071.
 PR 08-APR-1998; 98US-0081071.
 PR 09-APR-1998; 98US-0081195.
 PR 09-APR-1998; 98US-0081203.
 PR 09-APR-1998; 98US-0081229.
 PR 15-APR-1998; 98US-0081817.
 PR 15-APR-1998; 98US-0081838.
 PR 15-APR-1998; 98US-0081952.
 PR 15-APR-1998; 98US-0081955.
 PR 21-APR-1998; 98US-0082568.
 PR 21-APR-1998; 98US-0082569.
 PR 22-APR-1998; 98US-0082700.
 PR 22-APR-1998; 98US-0082704.
 PR 23-APR-1998; 98US-0082804.
 PR 23-APR-1998; 98US-0082767.
 PR 23-APR-1998; 98US-0082796.

Query Match 56.0%; Score 14; DB 22; Length 2405;
 Best Local Similarity 100.0%; Pred. No. 48;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

PR 27-APR-1998; 98US-0083336.
 PR 28-APR-1998; 98US-0083322.
 PR 29-APR-1998; 98US-0083392.
 PR 29-APR-1998; 98US-0083495.
 PR 29-APR-1998; 98US-0083496.
 PR 29-APR-1998; 98US-0083499.
 PR 29-APR-1998; 98US-0083500.
 PR 29-APR-1998; 98US-0083545.
 PR 29-APR-1998; 98US-0083554.
 PR 29-APR-1998; 98US-0083558.
 PR 29-APR-1998; 98US-0083559.
 PR 30-APR-1998; 98US-0083742.
 PR 05-MAY-1998; 98US-0084366.
 PR 06-MAY-1998; 98US-0084414.
 PR 06-MAY-1998; 98US-0084441.
 PR 07-MAY-1998; 98US-0084598.
 PR 07-MAY-1998; 98US-0084600.
 PR 07-MAY-1998; 98US-0084627.
 PR 07-MAY-1998; 98US-0084637.
 PR 07-MAY-1998; 98US-0084639.
 PR 07-MAY-1998; 98US-0084640.
 PR 07-MAY-1998; 98US-0084643.
 PR 13-MAY-1998; 98US-0085323.
 PR 13-MAY-1998; 98US-0085338.
 PR 13-MAY-1998; 98US-0085339.
 PR 15-MAY-1998; 98US-0085573.
 PR 15-MAY-1998; 98US-0085579.
 PR 15-MAY-1998; 98US-0085580.
 PR 15-MAY-1998; 98US-0085582.
 PR 15-MAY-1998; 98US-0085689.
 PR 15-MAY-1998; 98US-0085697.
 PR 15-MAY-1998; 98US-0085700.
 PR 15-MAY-1998; 98US-0085704.
 PR 18-MAY-1998; 98US-0086023.
 PR 22-MAY-1998; 98US-0086392.
 PR 22-MAY-1998; 98US-0086414.
 PR 22-MAY-1998; 98US-0086430.
 PR 22-MAY-1998; 98US-0086486.
 PR 28-MAY-1998; 98US-0087098.
 PR 28-MAY-1998; 98US-0087106.
 PR 28-MAY-1998; 98US-0087208.
 PR 30-JUL-1998; 98US-0094651.
 PR 11-SEP-1998; 98US-0100038.
 XX (GETH) GENENTECH INC.
 XX Wood WI, Goddard A, Gurney A, Yuan J, Baker KP, Chen J;
 PI P-PSDB; AA41699.
 DR WPI: 1999-551358/46.
 XX P-PSDB; AA41699.
 PT New secreted and transmembrane polypeptides and their polynucleotides,
 PT useful for treating blood coagulation disorders, cancers and cellular
 PT adhesion disorders -
 XX
 PS Claim 2; Fig 38; 530pp; English.
 XX
 CC The present invention describes secreted and transmembrane polypeptides
 CC and their polynucleotides. The nucleotide sequences are useful as
 CC sources of probes, primers, for chromosome mapping, and for generation
 CC of antisense sequences. They can also be used to create transgenic
 CC animals. The proteins can be used to treat a variety of diseases and
 CC disorders, depending on their function. Diseases that may be treated
 CC include blood coagulation disorders, cancers and cellular adhesion
 CC disorders. They may also be used to raise antibodies. AA433891 to
 CC AA23338, and AA41685 to AA41774 represent polynucleotides and
 CC polypeptide sequence given in the exemplification of the present
 CC invention.
 XX
 XX Sequence 2574 BP; 470 A; 775 C; 821 G; 508 T; 0 other;
 S0

Query Match

56.0%; Score 14; DB 20; Length 2574;

Best Local Similarity 100.0%; Pred. No. 48;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Caps 0;
 Oy 4 tctctgttgaag 17
 |||||
 Db 2167 TCTGCTGTTGAAG 2154
 RESULT 22
 AAC78481/c
 ID AAC78481 standard; cDNA; 2574 BP.
 XX
 AC AAC78481;
 XX
 XX 08-FEB-2001 (first entry)
 DT
 XX Human PRO703 (UNQ367) nucleotide sequence SEQ ID NO:101.
 DE
 XX Human; secreted protein; transmembrane protein; PRO; EST; cytosolic;
 KW expressed sequence tag; detection; cancer; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200053756-A2.
 XX
 PD 14-SEP-2000.
 XX
 PF 18-FEB-2000; 2000WO-US04341.
 XX
 PR 08-MAR-1999; 99WO-US05028.
 PR 12-MAR-1999; 99US-0123957.
 PR 29-MAR-1999; 99US-0126773.
 PR 21-APR-1999; 99US-0130232.
 PR 28-APR-1999; 99US-0131445.
 PR 14-MAY-1999; 99US-0134287.
 PR 23-JUN-1999; 99US-0141037.
 PR 26-JUL-1999; 99US-0145698.
 PR 29-OCT-1999; 99US-0162506.
 PR 30-NOV-1999; 99WO-US28313.
 PR 02-DEC-1999; 99WO-US28551.
 PR 02-DEC-1999; 99WO-US28565.
 PR 16-DEC-1999; 99WO-US30095.
 PR 30-DEC-1999; 99WO-US31243.
 PR 30-DEC-1999; 99WO-US31274.
 PR 05-JAN-2000; 2000WO-US00219.
 PR 06-JAN-2000; 2000WO-US00277.
 PR 06-JAN-2000; 2000WO-US00376.
 XX
 XX (GETH) GENENTECH INC.
 XX Ashkenazi AJ, Baker KP, Bolstein D, Desnoyers L, Eaton DL;
 PI Ferrara N, Fillaroff E, Fong S, Gao W, Gerber H, Gerritsen ME;
 PI Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ;
 PI Kiyavlin IU, Kuo SS, Napier MA, Pan J, Paoni NF, Roy MA;
 PI Shelton DL, Stewart TA, Tumas D, Williams PM, Wood WI;
 XX
 DR WPI: 2000-611443/58.
 DR P-PSDB; AAB44255.
 XX
 PT Novel PRO polypeptides and polynucleotides used in detection methods,
 PT to target bioactive molecules to specific cells, and to modulate
 PT cellular activities -
 XX
 PS Claim 2; Fig 38; 636pp; English.
 XX
 CC AAC78458 to AAC78599 represent polynucleotide and EST (expressed
 CC sequence tag) sequences which encode secreted or transmembrane PRO
 CC polypeptides. The PRO polynucleotides and polypeptides can be used for detecting
 CC activity. The polynucleotides and polypeptides can be used for detecting
 CC the presence of PRO polypeptides in samples, for linking bioactive
 CC molecules to cells and for modulating biological activities of cells,
 CC using the polypeptides for specific targeting. The polypeptide targeting
 CC can be used to kill the target cells, e.g. for the treatment of cancers.

CC The polypeptide pairs provide specific targeting of bioactive molecules
 CC to cells. AAC78600 to AAC78987 represent PCR primers and probes used in
 CC the isolation of the PRO polynucleotide sequences.

XX Sequence 2574 BP; 470 A; 775 C; 821 G; 508 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 2574;
 Best Local Similarity 100.0%; Pred. No. 48;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 tctgctgttgaag 17
 DB 2167 TCTGCTGTTGAAG 2154

RESULT 23
 AAC58239/C
 ID AAC58239 standard; cDNA: 2574 BP.

XX AAC58239;

DT 25-JAN-2001 (first entry)

XX Human PRO703 nucleotide sequence SEQ ID NO:28.

DE Human: tumour; diagnosis; neoplastic disease; identification; cancer;
 KM tumorigenesis; detection; neoplastic cell growth; proliferation;
 KM cytostatic; antiinflammatory; immunomodulatory; inflammatory disorder;
 KM immunological disorder; ss.

XX Homo sapiens.

PN W0200053754-A1.

XX 14-SEP-2000.

PD 06-JAN-2000; 2000MO-US00277.

XX 08-MAR-1999; 99MO-US05028.

PR 12-MAR-1999; 99US-0123957.

PR 29-MAR-1999; 99US-0126773.

PR 21-APR-1999; 99US-0130232.

PR 28-APR-1999; 99US-0131445.

PR 05-OCT-1999; 99MO-US23089.

PR 30-NOV-1999; 99MO-US28313.

PR 02-DEC-1999; 99MO-US28551.

PR 02-DEC-1999; 99MO-US28564.

PR 30-DEC-1999; 99MO-US31243.

PR 30-DEC-1999; 99MO-US31274.

XX (GETH) GENENTECH INC.

XX Baker KP, Desauvage FJ, Goddard A, Gurney AL, Klein RD, Roy MA;

XX Wood WI;

XX WPI: 2000-572269/53.

XX P-PSDB: AAB24054.

XX New isolated antibody for use in compositions and methods for the
 PT diagnosis and treatment of neoplastic cell growth and proliferation in
 PT mammals, including humans, and in monitoring tumor treatment -
 PS Claim 50; Fig 28; 195pp; English.

XX The present invention describes an isolated antibody (Ab) that binds to
 CC one of the human proteins (P) designated PRO213, PRO1330, PRO1449,
 CC PRO337, PRO324, PRO351, PRO615, PRO538, PRO3664, PRO618,
 CC PRO772, PRO703, PRO792 or PRO474. The Ab can be used in compositions
 CC and methods for the diagnosis and treatment of neoplastic cell growth
 CC and proliferation in mammals, including humans. Genes and polypeptides
 CC encoded by them, that are amplified in the genome of a tumor cell, can
 CC be identified and are useful targets for the treatment and prevention of

CC certain cancers and may be used to monitor tumor treatment. Compounds
 CC that inhibit the expression or activity of the identified polypeptides
 CC can be identified and used as antagonists. Benign or malignant tumours,
 CC inflammatory disorders and immunological disorders can be treated.
 CC AAC58123 to AAC58224 represent hybridisation probes and PCR primers used
 CC in the isolation of the human PRO sequences. AAC58225 to AAC58241 and
 CC AAB24041 to AAB24056 represent human PRO polynucleotide and protein
 CC sequences given in the exemplification of the present invention.

XX Sequence 2574 BP; 470 A; 775 C; 821 G; 508 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 2574;
 Best Local Similarity 100.0%; Pred. No. 48;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 tctgctgttgaag 17
 DB 2167 TCTGCTGTTGAAG 2154

RESULT 24
 AA25020/C

XX AA25020 standard; cDNA: 2669 BP.

XX AA25020;

DT 07-DEC-1999 (first entry)

XX Human H-beta 58 family CBALC05 nucleotide sequence.

DE Human: H-beta 58 family; CBALC05; cancer; AIDS; metabolic disorder;
 KM neurological disease; developmental abnormality; diagnosis;
 KM spontaneous abortion; ss.

XX Homo sapiens.

PN W09947663-A1.

XX 23-SEP-1999.

PD 18-MAR-1998; 98MO-CN00042.

PR 18-MAR-1998; 98MO-CN00042.

PR 18-MAR-1998; 98MO-CN00042.

PR 18-MAR-1998; 98MO-CN00042.

PR 18-MAR-1998; 98MO-CN00042.

XX (UYSH-) UNIV SHANGHAI SECOND MEDICAL.

XX Zhang Q, Kan L, He K, Shen Y;

XX WPI: 1999-571836/48.

XX P-PSDB: AAY41679.

XX CBALC05, a homologue of murine H-beta 58, and related polynucleotides

XX Claim 4; Page 9-11; 33pp; English.

XX The present sequence encodes a human H-beta 58 family protein designated

CC CBALC05. CBALC05 is homologous to the murine H-beta 58 protein. The

CC CBALC05 polynucleotide (1), vectors containing (1) and recombinant host

CC cells are useful for recombinant production of CBALC05. Both (1),

CC CBALC05 and antibodies against CBALC05 are useful as research reagents,

CC for screening assays and in diagnostic assays. Antagonists and agonists

CC of CBALC05 can be used to inhibit or enhance, respectively, the activity

CC of CBALC05 or expression of (1). Anti-CBALC05 antibodies and CBALC05

CC or its fragments can be used in vaccines. In particular, the proteins,

CC antibodies, agonists and antagonists can be used for treating, e.g.

CC cancer, AIDS, metabolic disorders, neurological disease, developmental

CC abnormalities and spontaneous abortions, related to both an excess and
 CC insufficient amounts of CBALC05.
 XX
 XX Sequence 2669 BP; 864 A; 419 C; 530 G; 856 T; 0 other;

Query Match 56.0%; Score 14; DB 20; Length 2669;
 Best Local Similarity 100.0%; Pred. No. 48;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 ctgctgtttgaagc 18
 |||||
 Db 949 CTGCTGTTTGAAGC 936

RESULT 25
 AAC98021/c
 ID AAC98021 standard; cDNA; 2675 BP.

XX AAC98021:
 AC
 XX 09-MAR-2001 (first entry)

DE Human colon cancer antigen nucleotide sequence SEQ ID NO:31.

XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
 KW identification; cystostatic; cardioactive; neuroprotective; vulnary;
 KW immunomodulatory; muscular; gynaecological; gastrointestinal;
 KW nephrotropic; antiinfective; antibacterial; gene therapy; wound;
 KW neural disorder; immune system disorder; muscular disorder;
 KW reproductive disorder; gastrointestinal disorder; renal disorder;
 KW infectious disease; cardiovascular disorder; ss.

XX Homo sapiens.
 OS
 XX WC000055351-A1.

XX 21-SEP-2000.

PD 08-MAR-2000; 2000MO-US05883.

XX 12-MAR-1999; 99US-0124270.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Ruben SM;

XX WPI: 2000-587534/55.

XX P-PSDB: AAB53264.

XX Colon cancer associated gene sequences, referred to as colon cancer
 PT antigens, useful for the treatment, prevention, and diagnosis of colon
 PT disorders such as colon cancer -

XX Claim 1; Page 490-491; 2104pp: English.

XX AAC97991 to AAC98763 encode the human colon cancer associated proteins,
 CC called human colon cancer antigens, given in AAB53234 to AAB54006. The
 CC human colon cancer antigens can have cystostatic, cardioactive, muscular;
 CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal,
 CC vulnary, nephrotropic, antiinfective and antibacterial activities, and
 CC can be used in gene therapy. The colon cancer antigen polynucleotides,
 CC proteins and antibodies to the proteins are useful for the prevention,
 CC treatment and diagnosis of colon disorders, such as colon cancer. The
 CC polynucleotides may be used in diagnostics and research, such as for
 CC chromosome identification, and as hybridisation probes. The proteins
 CC may also be used to prevent diseases such as neural disorders, immune
 CC system disorders, muscular disorders, reproductive disorders, infectious
 CC gastrointestinal disorders, wounds, renal disorders, infectious
 CC diseases, and cardiovascular disorders. AAC98764 to AAC98772 and
 CC AAB54007 represent sequences used in the exemplification of the present
 CC invention.

SQ Sequence 2675 BP; 856 A; 425 C; 532 G; 855 T; 7 other;

Query Match 56.0%; Score 14; DB 21; Length 2675;
 Best Local Similarity 100.0%; Pred. No. 48;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 ctgctgtttgaagc 18
 |||||
 Db 972 CTGCTGTTTGAAGC 959

RESULT 26
 AAV74359/c
 ID AAV74359 standard; DNA; 7778 BP.

XX AAV74359;

XX 16-MAR-1999 (first entry)

DE Staphylococcus aureus contig SEQ ID #48.

XX Computer readable medium; vaccine; S.aureus infection; immunodetection;
 KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
 KW skin infection; surgical wound infection; scalded skin syndrome;
 KW toxic shock syndrome; ds.

XX Staphylococcus aureus.

XX Key

XX misc_feature

XX 1681..1740

XX /tag= a

XX /note= "these bases represent a line of missing text in

XX the sequence listing in the specification. They

XX are included to maintain the nucleotide numbering

XX given in the specification for this DNA sequence"

XX misc_feature

XX 5281..5340

XX /tag= c

XX /note= "these bases represent a line of missing text in

XX the sequence listing in the specification. They

XX are included to maintain the nucleotide numbering

XX given in the specification for this DNA sequence"

XX misc_feature

XX 7081..7140

XX /tag= d

XX /note= "these bases represent a line of missing text in

XX the sequence listing in the specification. They

XX are included to maintain the nucleotide numbering

XX given in the specification for this DNA sequence"

XX EP786519-A2.

XX 30-JUL-1997.

XX 07-JAN-1997; 97EP-0100117.

XX 05-JAN-1996; 96US-0009861.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA;

XX Rosen CA;

XX WPI: 1997-374922/35.

XX Polynucleotide(s) and proteins derived from Staphylococcus aureus

XX stored on computer readable medium and used in the production of

XX anti-S.aureus vaccines

XX Claim 1: Page 387-391: 3271pp; English.

PS This sequence represents one of 5191 *Staphylococcus aureus* DNA sequences
 CC of the invention. The DNA sequences are recorded on a computer readable
 CC medium, preferably selected from a floppy or hard disk, random access
 CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
 CC the *S.aureus* DNA sequences allows putative functions to be assigned so
 CC that protein-encoding or regulatory regions of commercial, therapeutic or
 CC industrial importance can be obtained. Specifically, sequences which are
 CC likely to encode antigens have been identified and these polypeptides can
 CC be used in a vaccine composition against *S.aureus* infection. The
 CC polypeptides can also be used in a kit for the immunodetection of
 CC *S.aureus* in a sample. *S.aureus* is implicated in numerous human diseases,
 CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
 CC skin and surgical wound infections, scalded skin syndrome, toxic shock
 CC syndrome, etc. Organisms transformed with the DNA sequences can be used
 CC for recombinant production of the polypeptides. The new DNA sequences
 CC (and their fragments) are useful as primers or probes for isolating
 CC homologues of any of the *S.aureus* DNA sequences contained on the
 CC computer readable medium.

XX Sequence 7778 BP; 2771 A; 1071 C; 1452 G; 2239 T; 245 other;

Query Match 56.0%; Score 14; DB 18; Length 7778;
 Best Local Similarity 100.0%; Pred. No. 49;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 atctgctgttgaag 16
 |||||
 Db 4098 ATCTGCTGTTGAA 4085

RESULT 27
 AAQ06001

ID AAQ06001 standard; DNA: 12284 BP.

XX AC AAQ06001;

XX DT 16-JAN-1991 (first entry)

XX DE Sequence encoding protein characteristic of hog cholera virus (HCV).

XX KW Vaccine; parvovirus; swine influenza virus.

XX OS Hog cholera virus.

XX FH Key Location/Qualifiers

XX FT CDS 364..12057

XX FT /*tag= a

XX PN EP89034-A.

XX PD 26-SEP-1990.

XX PF 12-MAR-1990; 90EP-0200573.

XX PR 19-MAR-1989; 89EP-0104921.

XX PA 12-MAR-1990; 90EP-0200573.

XX PA (ALKU) AKZO NV.

XX PI Meyers G, Ruemenapf T, Thiel HJ;

XX DR WPI: 1990-291979/39.

XX DR P-PSDB; AAR06996.

XX PT New hog cholera virus vaccine and diagnostic - comprises nucleic

XX PS acid sequence of poly-peptide characteristic of hog cholera virus

CC Gene product may be used to provide a vaccine and Abs for diagnosis
 CC of hog cholera viral infection in pigs.
 XX Sequence 12284 BP; 3793 A; 2601 C; 3228 G; 2662 T; 0 other;

Query Match 56.0%; Score 14; DB 11; Length 12284;
 Best Local Similarity 100.0%; Pred. No. 49;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 tctgctgttgaag 17
 |||||
 Db 2253 tctgctgttgaag 2266

RESULT 28
 AAQ71308

ID AAQ71308 standard; DNA: 12284 BP.

XX AC AAQ71308;

XX DT 25-APR-1995 (first entry)

XX DE Sequence encoding 55 kilodalton protein of hog cholera virus.

XX KW Probe: detection; identification; diagnosis; hog cholera virus;

XX KM bovine viral diarrhoea virus; HCV; BVDV; ds.

XX OS Synthetic.

XX FH Key Location/Qualifiers

XX FT CDS 364..12060

XX FT /*tag= a

XX PN EP614979-A.

XX PD 14-SEP-1994.

XX PF 12-MAR-1990; 90EP-0200573.

XX PR 19-MAR-1989; 89EP-0104921.

XX PA (ALKU) AKZO NOBEL NV.

XX PI Meyers G, Ruemenapf T, Thiel H;

XX DR WPI: 1994-281207/35.

XX DR P-PSDB; AAR60543.

XX PT Hog cholera virus nucleic acid and polypeptide(s) - used to

XX PT develop prods. for use as vaccines and for diagnosis of hog

XX PS cholera virus infection

XX Claim 1: Figure 2; 63pp; English.

XX CC Antigenic fragments of the Hog Cholera Virus (HCV) 55 kDa

XX CC polypeptide may be used to induce virus neutralising antibodies.

XX CC The antigenic fragments; cells capable of producing them and

XX CC recombinant HCV may be used in vaccines to protect animals against

XX CC HCV infection. A nucleic acid sequence comprising a vector nucleic

XX CC acid molecule and encoding an antigenic fragment may be used for the

XX CC diagnosis of HCV infection and can discriminate HCV from bovine

XX CC viral diarrhoea virus. The antigenic fragments can also be used in

XX CC diagnostic methods to detect the presence of HCV.

XX Sequence 12284 BP; 3793 A; 2600 C; 3230 G; 2661 T; 0 other;

Query Match 56.0%; Score 14; DB 15; Length 12284;
 Best Local Similarity 100.0%; Pred. No. 49;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 tctgctgttgaag 17
 |||||||
 Db 2253 tctgctgttgaag 2266

RESULT 29
 AAT29591
 ID AAT29591 standard; cDNA; 12284 BP.
 XX
 AC AAT29591;
 XX
 DF 12-AUG-1996 (first entry)
 XX
 DE Hog cholera virus.
 XX
 KW Hog cholera virus; HCV; swine fever; vaccine; vector; diagnosis; ds.
 XX
 OS Hog cholera virus strain Alfort.
 XX
 PH Key
 PH Variation
 FT Location/Qualifiers
 FT 127
 FT /*tag= a
 FT /note= "replace(127, 'c')"
 FT CDS 364..12060
 FT /*tag= b
 FT /product= 435_KDa_protein
 FT 1522
 FT /*tag= c
 FT /note= "replace(1522, 'g'), alters codon 387 from
 FT ACC (Thr) to GCC (Ala)"
 FT 10989
 FT /*tag= d
 FT /note= "replace(10989, 't'), alters codon 3542
 FT from AGG (Arg) to AGT (Ser)"
 FT primer_bind
 FT complement (2587..2619)
 FT /*tag= e
 FT /note= "primer 1"
 FT complement (2842..2880)
 FT /*tag= f
 FT /note= "primer 2"
 XX
 PN EP713915-A1.
 XX
 PD 29-MAY-1996.
 XX
 PF 12-MAR-1990; 90EP-0200200.
 XX
 PR 19-MAR-1989; 89EP-0104921.
 XX
 PA (ALKU) AKZO NOBEL NV.
 XX
 PI Meyers G, Ruemenapf T, Thiel H;
 XX
 DR WPI; 1996-253042/26.
 DR P-PSDB; AAR95239.
 XX
 PT DNA encoding hog cholera virus polypeptide - part of hog cholera
 PT virus precursor poly:protein, useful in vaccines and in diagnostic
 PT tests.
 XX
 PS Claim 2; Page 12-30; 65pp; English.
 XX
 CC A cDNA sequence (AAT29591) derived from the genomic RNA of hog
 CC cholera virus (HCV) codes for a 435 kDa polypeptide (AAR95239)
 CC which, during HCV replication in an infected cell, is processed
 CC into the structural and non-structural proteins of the virus. The
 CC cDNA library produced from PK15 and 38A1D cells infected with HCV
 CC strain Alfort and screened with goat anti-HCV antiserum. The cDNA,
 CC or portions of it (see also AAT29594), can be used for prodn. of
 CC recombinant proteins useful as vaccines or diagnostics. It may
 CC also be incorporated into a vector, e.g. vaccinia virus, for use in
 CC vaccines, and primers based on the sequence (see also AAT29592-93)

CC are useful for diagnosis of hog cholera.
 XX
 XX Sequence 12284 BP; 3793 A; 2599 C; 3230 G; 2662 T; 0 other;
 SQ

Query Match 56.0%; Score 14; DB 17; Length 12284;
 Best Local Similarity 100.0%; Pred. No. 49;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 tctgctgttgaag 17
 |||||||
 Db 2253 tctgctgttgaag 2266

RESULT 30
 AAF28546
 ID AAF28546 standard; DNA; 63563 BP.
 XX
 AC AAF28546;
 XX
 DT 04-APR-2001 (first entry)
 XX
 DE Genomic fragment #33.
 XX
 KW Genomic library; bacteria; human upper airway; otitis media; sinusitis;
 KW bronchopulmonary; endocarditis; meningitis; ss.
 XX
 OS Moraxella catarrhalis.
 XX
 PN WO200078968-A2.
 XX
 PD 28-DEC-2000.
 XX
 PF 16-JUN-2000; 2000WO-US16649.
 XX
 PR 18-JUN-1999; 9905-0140121.
 XX
 PA (INCY-) INCYTE GENOMICS INC.
 XX
 PI Lagace RE, Patterson C, Berg KL;
 XX
 DR WPI; 2001-041427/05.
 XX
 PT Genomic library for identifying diagnostic and therapeutic
 PT compositions, and for identifying virulence factors, regulatory
 PT elements and drug targets, comprises Moraxella catarrhalis nucleic
 PT acids -
 XX
 XX Claim 1; Page 309-324; 545pp; English.
 PS
 XX
 CC The present invention relates to a Moraxella catarrhalis genomic library
 CC comprising of a combination of 41 nucleic acid molecules (see
 CC AAF28544-AAF28554). The library has a number of uses described in the
 CC specification e.g. is useful for identifying diagnostic and therapeutic
 CC compositions. M. catarrhalis (Branhamella catarrhalis) is a large
 CC aerobic, gram-negative diplococcus, normally found among the bacterial
 CC flora of human upper airways. M. catarrhalis is known to cause acute,
 CC localised infections such as otitis media, sinusitis and bronchopulmonary
 CC infection and life-threatening, systemic diseases including endocarditis
 CC and meningitis.
 XX
 XX Sequence 63563 BP; 18440 A; 12630 C; 11985 G; 20508 T; 0 other;
 SQ

Query Match 56.0%; Score 14; DB 22; Length 63563;
 Best Local Similarity 100.0%; Pred. No. 51;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 gatcgtcgttga 15
 |||||||
 Db 23739 gatcgtcgttga 23752

```

RESULT 31
AA094360/c
ID AA094360 standard; DNA; 17 BP.
XX
AC AA094360;
XX
DT 04-JUN-1996 (first entry)
XX
DE Septoria nodorum ITS primer JB434.
XX
KW Plant pathogen; fungus; Septoria nodorum; Septoria tritici; Fusarium;
KW Pseudocercospora herpotrichoides; Mycosphaerella filijensis; PCR;
KW Mycosphaerella muscicola; amplification; primer; ribosomal RNA gene;
KW internal transcribed region; strain; capture; colourimetric assay;
KW isolate; development; population; ss.
XX
OS Synthetic.
XX
PN W09529260-A2.
XX
PD 02-NOV-1995.
XX
PF 19-APR-1995; 95WO-US04712.
XX
PR 25-APR-1994; 94US-0233608.
XX
PA (CIBA ) CIBA GEIGY AG.
XX
PI Beck JJ, Ligon JM;
XX
DR WPI: 1995-383005/49.
XX
PT DNA encoding intervening transcribed sequence - used for detection
PS of plant fungal pathogens
XX
PS Claim 3; Page 14; 65pp; English.
XX
XX A novel method for the detection of plant pathogenic strains of fungi
CC e.g. Septoria nodorum, S.tritici, Pseudocercospora herpotrichoides,
CC Mycosphaerella filijensis, Muscicola or Fusarium spp, involves the PCR
CC amplification of sequences found in the internal transcribed region
CC (ITS) of the 18S, 5.8S and 28S ribosomal RNA genes by the primers
CC AA094359-93 and AA094357-72. These primers are derived from the ITS
CC sequences of these fungi (AA094359-93 and AA094357-72) and are strain
CC specific. The amplification products of the reactions using these
CC primers can be used with the capture primers AA094357-72 in
CC colourimetric assays. The primers and ITS DNAs can be used for the
CC detection of specific fungal pathogen isolates and in monitoring disease
CC development in plant populations.
XX
SQ Sequence 17 BP; 3 A; 4 C; 4 G; 6 T; 0 other;

Query Match 52.0%; Score 13; DB 16; Length 17;
Best Local Similarity 100.0%; Pred.No. 1.6e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 ttgaagcgcagca 24
DB 15 TTGAAGCGCAGCA 3
|||||
|

RESULT 32
AAV62508/c
ID AAV62508 standard; DNA; 17 BP.
XX
AC AAV62508;
XX
DT 17-DEC-1998 (first entry)
XX
DE Septoria nodorum species specific primer JB434.
XX
KW Internal transcribed spacer; ITS; ribosomal RNA; Fusarium avenaceum;

```

```

KW Fusarium culmorum; Fusarium graminearum; Fusarium moniliforme; plant;
KW Septoria avenae; Microdochium nivale; Fusarium poae; fungal pathogen;
KW PCR; nucleic acid detection; PCR primer; ss.
OS Synthetic.
OS Septoria nodorum.
XX US5814453-A.
XX 29-SEP-1998.
XX 02-JUL-1997; 97US-0887480.
XX 02-JUL-1997; 97US-0887480.
PR 19-APR-1995; 95WO-0504712.
PR 15-OCT-1996; 96US-0722187.
PA (NOVS ) NOVARTIS FINANCE CORP.
PI Beck JF:
XX WPI; 1998-541745/46.
XX DNA isolated from fungal RNA, and its internal transcribed spacer
PT sequence - used for detecting fungal pathogens in plant tissue
XX
XX Example 6; Column 16; 56pp; English.
XX
XX Sequences AAV62507 to AAV62566 represent species specific PCR primers
CC for various fungal isolates used for fungal detection in the course of
CC the invention. The primers are designed based on the internal
CC transcribed spacer (ITS) sequences of the various fungal species. The
CC invention provides a DNA molecule isolated from the ribosomal RNA gene
CC region of a fungal pathogen, where the DNA molecule consists of an ITS
CC sequence selected from ITS1 and ITS2 of Fusarium culmorum, Fusarium
CC graminearum, Fusarium moniliforme, Septoria avenae or Microdochium
CC nivale. A method for detecting F. graminearum, F. culmorum, F.
CC moniliforme, F. poae, F. avenaceum and M. nivale isolates is also
CC provided which comprises isolating DNA from a plant leaf infected with at
CC least one of the above pathogens and amplifying parts of the ITS sequence
CC of the pathogen(s) by PCR using specific primers from within these
CC sequences. The pathogen(s) are detected by visualising the amplified part
CC of the ITS sequence.
XX
XX Sequence 17 BP; 3 A; 4 C; 4 G; 6 T; 0 other:
SQ
Query Match 52.0%; Score 13; DB 19; Length 17;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0.
OY 12 ttgaagcagca 24
DB 15 TTGAAGCGCAGCA 3
|||||
RESULT 33
AA094361/c
1D AA094361 standard; DNA; 23 BP.
XX
XX AA094361;
XX
XX 04-JUN-1996 (first entry)
XX
XX Septoria nodorum ITS primer JB525.
XX
XX Plant pathogen; fungus; Septoria nodorum; Septoria tritici; Fusarium;
KW Pseudocercospora herpotrichoides; Mycosphaerella flajiensis; PCR;
KW Mycosphaerella muscicola; amplification; primer; ribosomal RNA gene;
KW internal transcribed region; strain; capture; colourimetric assay;
KW isolate; development; population; ss.
XX
XX Synthetic.

```

```

XX PN MO9529260-A2.
XX PD 02-NOV-1995.
XX PF 19-APR-1995; 95WO-US04712.
XX PR 25-APR-1994; 94US-0233608.
XX (CIBA ) CIBA GEIGY AG.
XX PI Beck JJ, Ligon JM;
XX WPI: 1995-383005/49.
XX PT DNA encoding intervening transcribed sequence - used for detection
XX of plant fungal pathogens
XX PS
XX Claim 3; Page 14; 65pp; English.
CC A novel method for the detection of plant pathogenic strains of fungi
CC e.g. Septoria nodorum, S. tritici, Pseudocercospora herpotrichoides,
CC Mycosphaerella filiformis, M. musicola or Fusarium spp. involves the PCR
CC amplification of sequences found in the internal transcribed region
CC (ITS) of the 18S, 5.8S and 28S ribosomal RNA genes by the primers
CC AA094359-93 and AAT05357-72. These primers are derived from the ITS
CC sequences of these fungi (AAT05394-R05404 and AA094398) and are strain
CC specific. The amplification products of the reactions using these
CC primers can be used with the capture primers AAT05378-93 in
CC colourimetric assays. The primers and ITS DNAs can be used for the
CC detection of specific fungal pathogen isolates and in monitoring disease
CC development in plant populations.
XX SQ Sequence 23 BP; 4 A; 6 C; 6 G; 7 T; 0 other:

Query Match 52.0%; Score 13; DB 16; Length 23;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 ttgaagcgacgca 24
   |||||||
Db 21 TTGAAGCGCAGCA 9

RESULT 34
AAV62509/c
ID AAV62509 standard; DNA: 23 BP.
XX
AC AAV62509;
XX
DT 17-DEC-1998 (first entry)
XX
DE Septoria nodorum species specific primer JB525.
XX
KM Internal transcribed spacer; ITS; ribosomal RNA; Fusarium avenaceum;
KM Fusarium culmorum; Fusarium graminearum; Fusarium moniliforme; plant;
KM Septoria avenae; Microdochium nivale; Fusarium poae; fungal pathogen;
KM PCR; nucleic acid detection; PCR primer; ss.
XX
OS Synthetic.
OS Septoria nodorum.
XX
PN US5814453-A.
XX
PD 29-SEP-1998.
XX
PF 02-JUL-1997; 97US-0867480.
XX
PR 02-JUL-1997; 97US-0867480.
PR 19-APR-1995; 95WO-US04712.
PR 15-OCT-1996; 96US-0722187.
XX

```

```

PA (NOV ) NOVARTIS FINANCE CORP.
XX
XX Beck JJ;
XX PI
XX WPI: 1998-541745/46.
XX
XX PT DNA isolated from fungal RNA, and its internal transcribed spacer
XX sequence - used for detecting fungal pathogens in plant tissue
XX
XX Example 6; Column 16; 56pp; English.
XX
XX Sequences AAV62507 to AAV62566 represent species specific PCR primers
XX for various fungal isolates used for fungal detection in the course of
XX the invention. The primers are designed based on the internal
XX transcribed spacer (ITS) sequences of the various fungal species. The
XX invention provides a DNA molecule isolated from the ribosomal RNA gene
XX region of a fungal pathogen, where the DNA molecule consists of an ITS
XX sequence selected from ITS1 and ITS2 of Fusarium culmorum, Fusarium
XX graminearum, Fusarium moniliforme, Septoria avenae or Microdochium
XX nivale. A method for detecting F. graminearum, F. culmorum, F.
XX moniliforme, F. poae, F. avenaceum and M. nivale isolates is also
XX provided which comprises isolating DNA from a plant leaf infected with at
XX least one of the above pathogens and amplifying parts of the ITS sequence
XX of the pathogen(s) by PCR using specific primers from within these
XX sequences. The pathogen(s) are detected by visualising the amplified part
XX of the ITS sequence.
XX
XX SQ Sequence 23 BP; 4 A; 6 C; 6 G; 7 T; 0 other:

Query Match 52.0%; Score 13; DB 19; Length 23;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 ttgaagcgacgca 24
   |||||||
Db 21 TTGAAGCGCAGCA 9

RESULT 35
AAA82189
ID AAA82189 standard; DNA: 128 BP.
XX
AC AAA82189;
XX
DT 04-DEC-2000 (first entry)
XX
DE N. meningitidis partial DNA sequence gnm_736 SEQ ID NO:736.
XX
KM Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
KM antigen; vaccine; diagnosis; infection; antibacterial; identification;
KM Meningococcus B; MenB; ds.
XX
OS Neisseria meningitidis.
XX
PN WO200022430-A2.
XX
PD 20-APR-2000.
XX
PF 08-OCT-1999; 99WO-US23573.
XX
PR 09-OCT-1998; 98US-0103794.
PR 30-APR-1999; 99US-0132068.
XX
XX (CHIR ) CHIRON CORP.
XX
XX Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;
XX PI Massignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;
XX PI Rappuoli R, Pizza M;
XX
XX WPI: 2000-318079/27.
XX
XX Isolated nucleotide sequences of Neisseria meningitidis which can be
PT

```

PT used in the diagnosis and treatment of N. meningitidis infection and
 PT other Neisserial infections, for example, N.gonorrhoea -

PS Claim 7; Page 1699; 1760pp; English.

CC The present invention describes methods of obtaining immunogenic
 CC proteins from Neisseria genomic sequences. AAA81453 to AAA82414
 CC represent specifically claimed Neisseria meningitidis genomic DNA
 CC sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent
 CC Neisseria DNA sequences and their corresponding proteins; AAA81254 to
 CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
 CC isolation of Neisseria meningitidis DNA sequences; and AAA81322 to
 CC AAA81452 represent Neisseria meningitidis MemB polynucleotide ORF
 CC sequences, which are all used in the exemplification of the present
 CC invention. The nucleic acid sequences, protein sequences, and antibodies
 CC against them, can be used in the manufacture of a composition. The
 CC composition can be used as a medicament (or in the manufacture of a
 CC medicament) for treating, preventing or diagnosing infection due to
 CC Neisserial bacteria. For example, some of the identified proteins could
 CC be components of vaccines against Meningococcus B; against all serotypes;
 CC and/or against all pathogenic Neisseriae. Identification of sequences
 CC from the bacterium will also facilitate production of biological probes,
 CC particularly organism-specific probes. Attempts to make efficacious
 CC Meningococcus B vaccines have failed mainly due to antigen tolerance.
 CC Multivalent vaccines have also been tried but none have successfully
 CC overcome antigenic variability. The provision of further, complete
 CC sequences may provide an opportunity to identify secreted or surface
 CC exposed proteins that may be presumed targets for the immune system and
 CC which are not antigenically variable or at least more conserved than
 CC other more variable regions.

CC Sequence 128 BP; 32 A; 35 C; 27 G; 34 T; 0 other;

SO

Query Match 52.0%; Score 13; DB 21; Length 128;
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 ctgctgttgaag 17
 |||||

Db 101 ctgctgttgaag 113

RESULT 36

AACT12134

ID AACT12134 standard; cDNA: 157 BP.

XX AACT12134;

DT 06-OCT-2000 (first entry)

DE Human secreted protein 5' EST, SEQ ID NO: 16209.

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;

KW gene therapy; chromosome mapping; ss.

OS Homo sapiens.

PN EP1033401-A2.

PD 06-SEP-2000.

PF 21-FEB-2000; 2000EP-0200610.

PR 26-FEB-1999; 99US-0122487.

PA (GEST) GENSET.

PI Dumas Milne Edwards J, Duclert A, Giordano J;

DR WPI; 2000-500381/45.

PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for

PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -

PS Claim 1; SEQ ID 16209; 71bp + CD-ROM; English.

CC The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
 CC identified within the present sequence. The 5' ESTs were prepared from
 CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA
 CC libraries. Such ESTs are not well suited for isolating cDNA sequences
 CC derived from the 5' ends of mRNAs and even in those cases where longer
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
 CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
 CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
 CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
 CC They are used to obtain upstream regulatory sequences and to design
 CC expression and secretion vectors.

CC Sequence 157 BP; 24 A; 31 C; 76 G; 25 T; 1 other;

SO

Query Match 52.0%; Score 13; DB 21; Length 157;
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 gctgttgaagc 19
 |||||

Db 89 gctgttgaagc 101

RESULT 37

AACT1863

ID AACT1863 standard; cDNA: 192 BP.

XX AACT1863;

DT 06-OCT-2000 (first entry)

DE Human secreted protein 5' EST, SEQ ID NO: 15938.

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;

KW gene therapy; chromosome mapping; ss.

OS Homo sapiens.

PN EP1033401-A2.

PD 06-SEP-2000.

PF 21-FEB-2000; 2000EP-0200610.

PR 26-FEB-1999; 99US-0122487.

PA (GEST) GENSET.

PI Dumas Milne Edwards J, Duclert A, Giordano J;

DR WPI; 2000-500381/45.

PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -

PS Claim 1; SEQ ID 15938; 71bp + CD-ROM; English.

CC The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
 CC identified within the present sequence. The 5' ESTs were prepared from
 CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
 CC sequences usually correspond mainly to the 3' untranslated region (UTR)
 CC of the mRNA because they are often obtained from oligo-dT primed cDNA

CC Libraries. Such ESTs are not well suited for isolating cDNA sequences
 CC derived from the 5' ends of mRNAs and even in those cases where longer
 CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
 CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
 CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
 CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
 CC They are used to obtain upstream regulatory sequences and to design
 CC expression and secretion vectors.

SO Sequence 192 BP; 59 A; 48 C; 42 G; 43 T; 0 other;

Query Match 52.0%; Score 13; DB 21; Length 192;

Best Local Similarity 100.0%; Pred. No. 1.6e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 atctgcctgttga 15
 |||||
 Db 137 atctgcctgttga 149

RESULT 38

AA012455 standard; DNA: 251 BP.

AA012455;

30-MAR-1999 (first entry)

Human diallelic polymorphic DNA fragment WI-19034.

DE Polymorphism; diallelic; human; forensic; paternity testing; disease;
 KM detection; phenotypic typing; characteristic; infection; hereditary;
 KM autoimmune disease; cancer; inflammation; drug; therapy; medication;
 KM treatment; marker; ss.

OS Homo sapiens.

PN W09820165-A2.

PD 14-MAY-1998.

PF 05-NOV-1997; 97MO-US20313.

PR 06-NOV-1996; 96US-0030455.

PA (WHEED) WHITEHEAD INST BIOMEDICAL RES.

PI Hudson T, Lander ES, Wang D;

DR WPI: 1998-286974/25.

PT New isolated nucleic acid segments from the human genome - used for
 PT determining polymorphic forms for use in e.g. forensics; paternity
 PT testing or phenotypic typing for disease

PS Claim 1; Page 256; 310pp; English.

XX AAX10269-X12937 are human DNA fragments which contain diallelic
 CC polymorphic markers which have been isolated using the primers
 CC represented in AAX09121-X10268. The base occupying the polymorphic site
 CC is indicated by the appropriate IUPAC-IUB ambiguity code. These fragments
 CC can be used in methods for determining polymorphic forms in an individual
 CC for use in e.g. forensics, paternity testing or for phenotypic typing for
 CC diseases such as agammaglobulinemia, diabetes insipidus, Lesch-Nyman
 CC syndrome, muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease,
 CC familial hypercholesterolemia, polycystic kidney disease, hereditary
 CC spherocytosis, von Willebrand's disease, tubercular sclerosis, Ehlers-Danlos
 CC haemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos
 CC syndrome, osteogenesis imperfecta, acute intermittent porphyria,
 CC autoimmune diseases, inflammation, cancer, diseases of the nervous
 CC system, infection by pathogenic microorganisms, and characteristics such
 CC as longevity, appearance (e.g. baldness, obesity), strength, speed.

CC endurance, fertility, and susceptibility or receptivity to particular
 CC drugs or therapeutic treatments. The isolated polymorphic nucleic acid
 CC segments can also be used to produce medicaments for the treatment or
 CC prophylaxis of such diseases.

SO Sequence 251 BP; 72 A; 45 C; 46 G; 87 T; 1 other;

Query Match 52.0%; Score 13; DB 19; Length 251;

Best Local Similarity 100.0%; Pred. No. 1.6e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 tctgcctgttga 16
 |||||
 Db 100 tctgcctgttga 112

RESULT 39

AAF09435/c standard; cDNA: 299 BP.

AAF09435;

13-MAR-2001 (first entry)

Fusarium venenatum EST SEQ ID NO:1958.

DE Multiple gene expression; filamentous fungal cell; EST;
 KM expressed sequence tag; Fusarium venenatum; Aspergillus niger;
 KM Aspergillus oryzae; Trichoderma reesei; identification; recombination;
 KM culture condition; environmental stress; spore morphogenesis;
 KM metabolic pathway engineering; catabolic pathway engineering; ss.

OS Fusarium venenatum.

PN W0200056762-A2.

PD 28-SEP-2000.

PF 22-MAR-2000; 2000WO-US07781.

PR 22-MAR-1999; 99US-0273623.

PA (NOVO) NOVO NORDISK BIOTECH INC.

PA (NOVO) NOVO NORDISK AS.

PI Berka RM, Rey MM, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;

DR WPI: 2000-594572/56.

PT Monitoring differential expression of genes in filamentous fungal cells
 PT uses fluorescence-labeled nucleic acids isolated from the cells and a
 PT substrate of expressed sequence tags -

PS Claim 86; Page 433; 3161pp; English.

XX The present invention describes a method for monitoring differential
 CC expression of genes in a first filamentous fungal (FF) cell relative to
 CC expression of the same genes in one or more second filamentous fungal
 CC cells. The method uses fluorescence-labeled nucleic acids isolated from
 CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
 CC are used in the methods for monitoring differential expression of the
 CC same genes in one or more second filamentous fungal cells. Monitoring
 CC the global expression of genes from FF cells allows the production
 CC potential of the microorganisms to be improved. New genes may be
 CC discovered, possible functions of unknown open reading frames can be
 CC identified and gene copy number variation and stability can be
 CC monitored. The expression of genes can be used to study how FF cells
 CC adapt to changes in culture conditions, environmental stress, spore
 CC morphogenesis, recombination, metabolic or catabolic pathway
 CC engineering. Using ESTs provides several advantages over genomic or
 CC random cDNA clones including elimination of redundancy as one spot on an

CC array equals one gene or open reading frame, and organisation of the
 CC microarrays based on function of the gene products to facilitate
 CC analysis of the results. AAF07478 to AAF11247 represents ESTs from
 CC Fuserium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
 CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
 CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
 CC all specifically claimed in the present invention.
 SQ Sequence 299 BP; 90 A; 79 C; 74 G; 56 T; 0 other;

Query Match 52.0%; Score 13; DB 21; Length 299;
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 tttagagcgacgc 23
 |||||
 DB 221 TTGAGCGCAGC 209

RESULT 40
 AAA01371/c
 ID AAA01371 standard; cDNA; 300 BP.
 AC AAA01371;

19-MAY-2000 (first entry)

Human colon cancer cell line polynucleotide sequence SEQ ID NO:1362.

Human; colon cancer; tumour; diagnosis; gene expression product;
 probe; detection; cancerous state; metastasis; identification;
 breast cancer; oestrogen receptor-positive breast cancer; therapy;
 oestrogen receptor-negative breast cancer; lung cancer; ss.

Homo sapiens.

WO958675-A2.

18-NOV-1999.

13-MAY-1999; 99WO-US10602.

14-MAY-1998; 98US-0085426.

15-MAY-1998; 98US-0085537.

21-OCT-1998; 98US-0085696.

27-OCT-1998; 98US-0105234.

(CHIR) CHIRON CORP.

(HYSE-) HYSEQ INC.

Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;

Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;

Lamson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;

Leshkowitz D, Kita D, Garcia V, Jones LM, Stache-Grain B;

WPI: 2000-126369/11.

Polynucleotide library used to determine cancerous states of mammalian

cells.

Claim 1; Page 509; 1097pp; English.

AAA00010 to AAA02716 represent polynucleotides isolated from cDNA
 libraries constructed from human colon cancer cell lines. The present
 invention also describes a method of detecting differentially expressed
 genes correlated with a cancerous state of a mammalian cell, comprising
 detecting at least one differentially expressed gene product in a test
 sample derived from a cell suspected of being cancerous, where detection
 of the differentially expressed gene product is correlated with a
 cancerous state of the cell from which the test sample was derived.
 The polynucleotides sequences can be used in a method for detecting

CC differentially expressed genes correlated with a cancerous state of a
 CC mammalian cell. The polynucleotides can also be used as probes for
 CC detecting and mapping related genes. They can be used in diagnosis and
 CC prognosis of diseases and disorders (e.g. identification of
 CC pre-metastatic or metastatic cancerous states, stages of cancer, or
 CC responsiveness of cancer to therapy). This is particularly for breast
 CC cancer, oestrogen receptor-positive breast cancer, oestrogen receptor-
 CC negative breast cancer, lung cancer, and colon cancer.
 SQ Sequence 300 BP; 86 A; 59 C; 57 G; 98 T; 0 other;

Query Match 52.0%; Score 13; DB 21; Length 300;
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 ctgctgtttgaag 17
 |||||
 DB 198 CTGCTGTTGAG 186

RESULT 41
 AAV01888/c
 ID AAV01888 standard; cDNA; 384 BP.
 AC AAV01888;

20-APR-1998 (first entry)

Human OTK27 gene.

Human; foetal brain cDNA library; GDP dissociation stimulating protein;
 brain specific nucleosome assembly protein; diagnosis; therapy;
 skeletal muscle specific ubiquitin conjugating enzyme; TMP-2; NPIK;
 nel-related protein type 1; nel-related type 2; hereditary disease;
 cancer; OTK27; ss.

Homo sapiens.

EP796913-A2.

24-SEP-1997.

19-MAR-1997; 97EP-0104842.

05-MAR-1997; 97JP-0069163.

19-MAR-1996; 96JP-0063410.

(SANA) OTSUKA PHARM CO LTD.

Fujiwara T, Horie M, Watanabe T;

WPI: 1997-459830/43.

P-PSDB; AAW37503.

Novel human genes, e.g. brain-specific nucleosome assembly protein -

useful for diagnosis or therapy of hereditary disease and cancer

Example 3; Page 35; 123pp; English.

The present sequence encodes a OTK27 isolated from a human foetal brain
 cDNA library. The nucleotide or amino acid sequences are useful for
 in-vitro diagnosis of hereditary diseases and cancer and for preparation
 of pharmaceuticals.

Sequence 384 BP; 94 A; 111 C; 98 G; 81 T; 0 other;

Query Match 52.0%; Score 13; DB 18; Length 384;
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ggcctcgtctt 13
 |||||
 Db 349 GGATCTGCTGTT 337

RESULT 42

AAf65138/c
 ID AAF65138 standard; cDNA: 391 BP.

AC AAF65138;

DT 09-APR-2001 (first entry)

DE Novel human polynucleotide, SEQ ID NO: 894.

KW Human; cytostatic; gene therapy; colon cancer; prostate cancer;
 breast cancer; lung cancer; cancer detection; ss.

OS Homo sapiens.

PN WO200102568-A2.

PD 11-JAN-2001.

PF 30-JUN-2000; 2000WO-US18374.

PR 02-JUL-1999; 99US-0142310.

PR 02-JUL-1999; 99US-0142311.

PA (CHIR) CHIRON CORP.
 (HYSE-) HYSEQ INC.

PI Williams LT, Escobedo J, Innis MA, Garcia PD, Klinger J, Kassam A;
 Reinhard C, Randazzo F, Kennedy GC, Pot D, Lamson G, Drmanac R;
 Cirenjakov R, Drmanac S, Dickson M, Labat I, Leshkowitz D;
 Kita D, Garcia V, Jones LW, Strache-Crain B;

DR WPI: 2001-091805/10.

PT Library of polynucleotides for diagnosing a cancerous state of a
 mammalian cell and detecting cancer, particularly of the colon or
 prostate, comprises 3351 human polynucleotide sequences -

PS Claim 9; Page 672-673; 1046pp; English.

CC The present sequence is one of 3351 sequences in a library of human
 CC polynucleotides. The library is used to detect differentially expressed
 CC genes correlated with a cancerous state of a mammalian cell and can
 CC detect colon, prostate, breast and lung cancer. The library can be used
 CC to produce probes for detection of mRNA and to produce additional copies
 CC of the polynucleotides. The probes can be used for chromosome mapping of
 CC the polynucleotide and for detection of transcription levels. Ribozymes
 CC or antisense oligonucleotides can be generated. The polynucleotides and
 CC their gene products are used as genetic or biochemical markers (e.g. in
 CC blood or tissues) that will detect the earliest changes along the
 CC carcinogenesis pathway and/or monitor the efficacy of therapies and
 CC preventive interventions. The polynucleotides, polypeptides and
 CC antibodies against them can be used in pharmaceutical compositions to
 CC treat the cancers and proliferative disorders such as neoplasia,
 CC dysplasia and hyperplasia.

SO Sequence 391 BP; 73 A; 118 C; 132 G; 68 T; 0 other;

Query Match 52.0%; Score 13; DB 22; Length 391;
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 ttgaagcgagca 24
 |||||
 Db 168 TTGAAGCGAGCA 156

RESULT 43

AAf66457/c
 ID AAF66457 standard; cDNA: 398 BP.

AC AAF66457;

DT 09-APR-2001 (first entry)

DE Novel human polynucleotide, SEQ ID NO: 2213.

KW Human; cytostatic; gene therapy; colon cancer; prostate cancer;
 breast cancer; lung cancer; cancer detection; ss.

OS Homo sapiens.

PN WO200102568-A2.

PD 11-JAN-2001.

PF 30-JUN-2000; 2000WO-US18374.

PR 02-JUL-1999; 99US-0142310.

PR 02-JUL-1999; 99US-0142311.

PA (CHIR) CHIRON CORP.
 (HYSE-) HYSEQ INC.

PI Williams LT, Escobedo J, Innis MA, Garcia PD, Klinger J, Kassam A;
 Reinhard C, Randazzo F, Kennedy GC, Pot D, Lamson G, Drmanac R;
 Cirenjakov R, Drmanac S, Dickson M, Labat I, Leshkowitz D;
 Kita D, Garcia V, Jones LW, Strache-Crain B;

DR WPI: 2001-091805/10.

PT Library of polynucleotides for diagnosing a cancerous state of a
 mammalian cell and detecting cancer, particularly of the colon or
 prostate, comprises 3351 human polynucleotide sequences -

PS Claim 9; Page 867; 1046pp; English.

CC The present sequence is one of 3351 sequences in a library of human
 CC polynucleotides. The library is used to detect differentially expressed
 CC genes correlated with a cancerous state of a mammalian cell and can
 CC detect colon, prostate, breast and lung cancer. The library can be used
 CC to produce probes for detection of mRNA and to produce additional copies
 CC of the polynucleotides. The probes can be used for chromosome mapping of
 CC the polynucleotide and for detection of transcription levels. Ribozymes
 CC or antisense oligonucleotides can be generated. The polynucleotides and
 CC their gene products are used as genetic or biochemical markers (e.g. in
 CC blood or tissues) that will detect the earliest changes along the
 CC carcinogenesis pathway and/or monitor the efficacy of therapies and
 CC preventive interventions. The polynucleotides, polypeptides and
 CC antibodies against them can be used in pharmaceutical compositions to
 CC treat the cancers and proliferative disorders such as neoplasia,
 CC dysplasia and hyperplasia.

SO Sequence 398 BP; 109 A; 76 C; 79 G; 134 T; 0 other;

Query Match 52.0%; Score 13; DB 22; Length 398;
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 ctgcgtttgaag 17
 |||||
 Db 254 CTGCTTTGAAG 242

Result	44
ID	AAA56847/c
AA	AAA56847 standard; DNA: 400 BP.
AC	AAA56847;
AD	
AE	20-OCT-2000 (first entry)
AF	
AG	Mycoplasma hyopneumoniae genomic clone pAD983.
AH	
AI	Mycoplasma hyopneumoniae; vaccine; antigen; antimicrobial;
AJ	enzootic pneumonia; ds.
AK	
AL	Mycoplasma hyopneumoniae.
AM	
AN	Key
AO	Location/Qualifiers
AP	1..400
AQ	CDS
AR	FT
AS	FT
AT	FT
AV	FT
AW	FT
AX	FT
AY	FT
AZ	FT
BA	FT
BB	FT
BC	FT
BD	FT
BE	FT
BF	FT
BG	FT
BH	FT
BI	FT
BJ	FT
BK	FT
BL	FT
BM	FT
BN	FT
BO	FT
BP	FT
BQ	FT
BR	FT
BS	FT
BT	FT
BU	FT
BV	FT
BW	FT
BX	FT
BY	FT
BZ	FT
CA	FT
CB	FT
CC	FT
CD	FT
CE	FT
CF	FT
CG	FT
CH	FT
CI	FT
CJ	FT
CK	FT
CL	FT
CM	FT
CN	FT
CO	FT
CP	FT
CQ	FT
CR	FT
CS	FT
CT	FT
CU	FT
CV	FT
CW	FT
CX	FT
CY	FT
CA	FT
CB	FT
CC	FT
CD	FT
CE	FT
CF	FT
CG	FT
CH	FT
CI	FT
CJ	FT
CK	FT
CL	FT
CM	FT
CN	FT
CO	FT
CP	FT
CQ	FT
CR	FT
CS	FT
CT	FT
CU	FT
CV	FT
CW	FT
CX	FT
CY	FT
CA	FT
CB	FT
CC	FT
CD	FT
CE	FT
CF	FT
CG	FT
CH	FT
CI	FT
CJ	FT
CK	FT
CL	FT
CM	FT
CN	FT
CO	FT
CP	FT
CQ	FT
CR	FT
CS	FT
CT	FT
CU	FT
CV	FT
CW	FT
CX	FT
CY	FT
CA	FT
CB	FT
CC	FT
CD	FT
CE	FT
CF	FT
CG	FT
CH	FT
CI	FT
CJ	FT
CK	FT
CL	FT
CM	FT
CN	FT
CO	FT
CP	FT
CQ	FT
CR	FT
CS	FT
CT	FT
CU	FT
CV	FT
CW	FT
CX	FT
CY	FT
CA	FT
CB	FT
CC	FT
CD	FT
CE	FT
CF	FT
CG	FT
CH	FT
CI	FT
CJ	FT
CK	FT
CL	FT
CM	FT
CN	FT
CO	FT
CP	FT
CQ	FT
CR	FT
CS	FT
CT	FT
CU	FT
CV	FT
CW	FT
CX	FT
CY	FT
CA	FT
CB	FT
CC	FT
CD	FT
CE	FT
CF	FT
CG	FT
CH	FT
CI	FT
CJ	FT
CK	FT
CL	FT
CM	FT
CN	FT</

AAx2975/c	AAx2975 standard; cDNA; 423 BP.
ID	AAx2975; standard; cDNA; 423 BP.
XX	
AC	AAx2975;
XX	
DT	06-JUL-1999 (first entry)
XX	
DE	Fragment of human placental STAT3 coding sequence.
XX	
KW	Placenta; isoform; human; STAT3; intracellular; transcription factor;
KW	Signal Transducer and Activator of Transcription; allele; growth arrest;
KW	hepatic acute-phase protein; monocytic cell; myeloma; autoimmune disease;
XX	inflammation; ss.
XX	
OS	Homo sapiens.
XX	
PN	EP06953-A1.
XX	
PD	07-APR-1999.
XX	
PF	16-SEP-1997; 97EP-0116061.
XX	
PR	16-SEP-1997; 97EP-0116061.
XX	
PA	(ISTF) ARS APPLIED RES SYSTEMS HOLDING NV.
XX	
PI	Della Pietra L, Serlupi-Crescenzi O;
DR	WPI: 1999-207107/18.
XX	
DR	P-PSDB: AAY07239.
XX	
PT	New allelic variant of human STAT3 useful in treating autoimmune or
XX	inflammatory diseases
XX	
PS	Disclosure; Page 8-9; 22pp; English.
XX	
CC	This sequence represents a fragment of the coding sequence for the
CC	placental isoform of human Signal Transducer and Activator of
CC	Transcription (STAT3) intracellular transcription factor (Akira et al.,
CC	Cell 77, 63-71 (1994)). The invention relates to isolation of allelic
CC	variants of the placental hSTAT3 sequence. hSTAT3 plays a role in the
CC	upregulation of hepatic acute-phase proteins, growth arrest of monocytic
CC	cells and in the survival of myeloma cells and so may be used to treat
CC	or diagnose autoimmune or inflammatory diseases.
XX	
SQ	Sequence 423 BP; 115 A; 106 C; 109 G; 93 T; 0 other;
XX	
QY	Query Match 52.0%; Score 13; DB 20; Length 423;
XX	Best Local Similarity 100.0%; Pred. No. 1.7e+02;
DB	Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0
142	TGAAGCGCAGCAG 130
13	tgaagcgccagcag 25

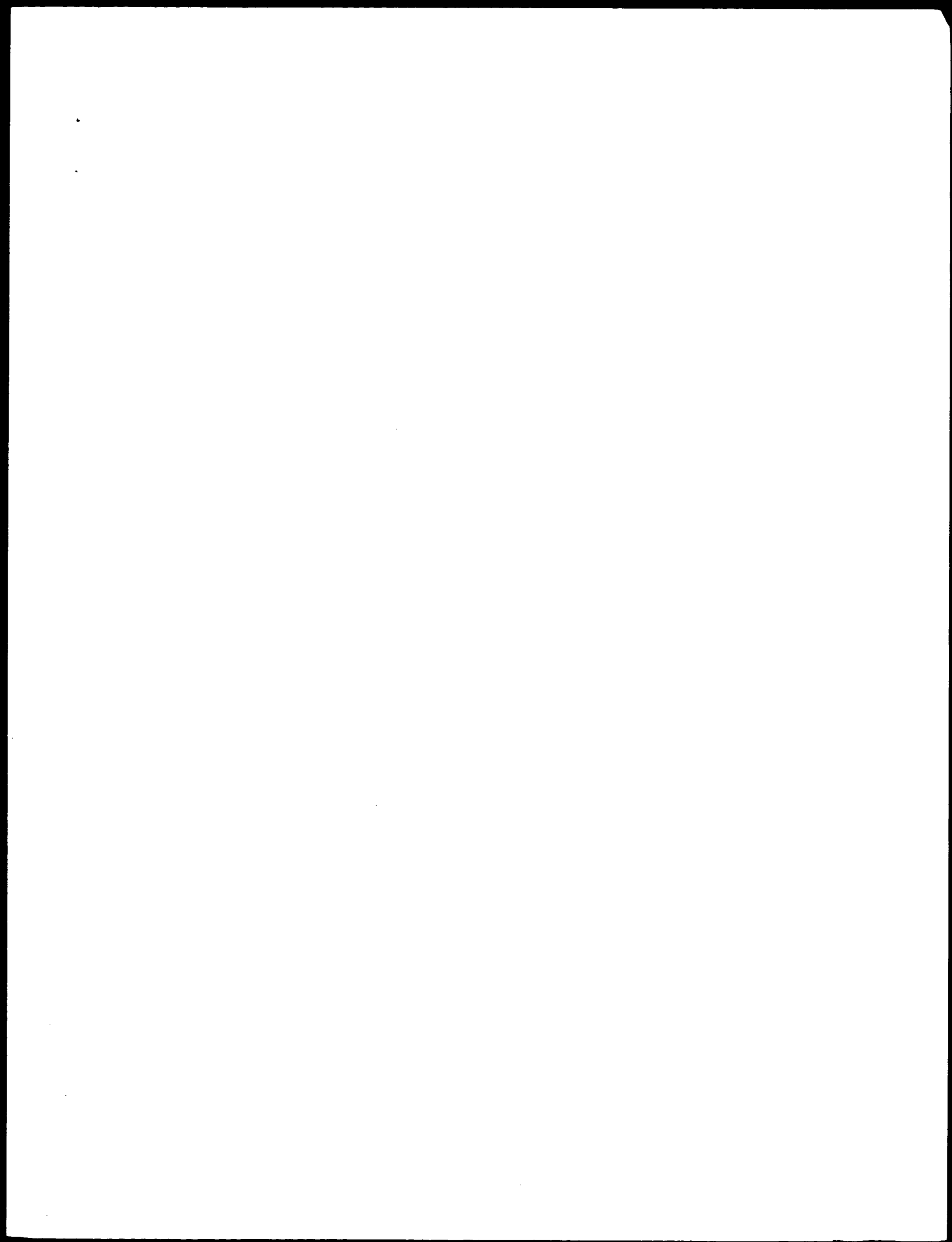
Search completed: October 9, 2001, 15:52:11
Job time: 15092 sec

RESULT 45

Wed Oct 10 07:45:49 2001

us-09-396-196f-4.oli.rng

Page 25



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 15:55:26 ; Search time 218.82 Seconds
(without alignments)
21.629 Million cell updates/sec

Title: US-09-396-196f-4

Perfect score: 25

Sequence: 1 ggatctgctgttgaagcagcagcag 25

Scoring table: OLIGO-NUC

Searched: 324599 seqs, 9465562 residues

Word size : 9

Total number of hits satisfying chosen parameters: 11319

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Issued Patents NA:*

1: /cgn2_6/prodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/prodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/prodata/2/ina/5A_COMB.seq:*
4: /cgn2_6/prodata/2/ina/5B_COMB.seq:*
5: /cgn2_6/prodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/prodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	100.0	1041	2	US-08-401-068-7
2	25	100.0	1041	2	US-08-846-338-7
3	25	100.0	5872	3	US-08-411-768B-1
4	25	100.0	5872	3	US-08-411-768B-6
5	14	56.0	12284	2	US-08-876-991-1
6	14	56.0	12284	2	US-09-059-853-1
7	13	52.0	17	1	US-08-233-608-8
8	13	52.0	17	1	US-08-887-480-8
9	13	52.0	17	1	US-08-722-187-8
10	13	52.0	17	5	PCT-US95-04712-8
11	13	52.0	23	1	US-08-233-608-9
12	13	52.0	23	1	US-08-887-480-9
13	13	52.0	23	1	US-08-722-187-9
14	13	52.0	23	5	PCT-US95-04712-9
15	13	52.0	27	1	US-08-150-331-31
16	13	52.0	384	2	US-08-820-170A-8
17	13	52.0	384	3	US-09-055-699-8
18	13	52.0	384	4	US-09-273-565-8
19	13	52.0	383	1	US-08-233-608-2
20	13	52.0	583	1	US-08-887-480-2
21	13	52.0	583	2	US-08-722-187-2
22	13	52.0	583	5	PCT-US95-04712-2
23	13	52.0	749	4	US-08-998-416-727
24	13	52.0	1161	1	US-08-153-848-31
25	13	52.0	1161	5	PCT-US93-11153-31
26	13	52.0	1242	1	US-08-252-966B-13
27	13	52.0	1425	1	US-08-353-550-7

C 28	13	52.0	1425	2	US-08-551-687-7	Sequence 7, Appl1
C 29	13	52.0	1493	2	US-08-820-170A-9	Sequence 9, Appl1
C 30	13	52.0	1493	3	US-09-055-699-9	Sequence 9, Appl1
C 31	13	52.0	1493	4	US-09-273-565-9	Sequence 9, Appl1
C 32	13	52.0	1712	1	US-08-171-299B-1	Sequence 1, Appl1
C 33	13	52.0	1743	3	US-09-221-235-9	Sequence 9, Appl1
C 34	13	52.0	1743	3	US-09-221-928-9	Sequence 9, Appl1
C 35	13	52.0	1743	3	US-09-221-577-9	Sequence 9, Appl1
C 36	13	52.0	1743	3	US-09-221-236-9	Sequence 9, Appl1
C 37	13	52.0	1743	3	US-09-221-416-9	Sequence 9, Appl1
C 38	13	52.0	1743	4	US-09-163-115-9	Sequence 9, Appl1
C 39	13	52.0	1743	4	US-09-163-528-9	Sequence 9, Appl1
C 40	13	52.0	1743	4	US-09-593-553-9	Sequence 9, Appl1
C 41	13	52.0	1743	4	US-09-221-237-9	Sequence 9, Appl1
C 42	13	52.0	1743	4	US-08-361-611-3	Sequence 3, Appl1
C 43	13	52.0	1835	1	US-08-565-655-3	Sequence 3, Appl1
C 44	13	52.0	1835	1	US-08-565-655-3	Sequence 3, Appl1
C 45	13	52.0	1835	2	US-08-946-967-3	Sequence 3, Appl1

ALIGNMENTS

RESULT 1
US-08-401-068-7
Sequence 7, Application US/08401068
Patent No. 5859335
GENERAL INFORMATION:
APPLICANT: Patton, David
TITLE OF INVENTION: Enhanced Biotin Biosynthesis in Plant Tissue
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/401,068
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/351,970
FILING DATE: 08-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
TELEPHONE: 919-541-8614
TELEFAX: 919-541-8689
INFORMATION FOR SEQ. ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1041 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1038
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /product= "biotin synthase"
OTHER INFORMATION: /evidence= EXPERIMENTAL
US-08-401-068-7

Query Match 100.0%; Score 25; DB 2; Length 1041;
Best Local Similarity 100.0%; Pred. No. 9.6e-06;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggatctgctgtttgaagcgcagcag 25
|||||
Db 63 GGATCTGCTGTTTGAAGCGCACAG 87

RESULT 2

US-08-846-338-7

Sequence 7, Application US/08846338

Patent No. 5869719

GENERAL INFORMATION:

APPLICANT: Patton, David

TITLE OF INVENTION: Transgenic Plants Having Increased Biotin Content

NUMBER OF SEQUENCES: 20

CORRESPONDENCE ADDRESS:

ADDRESSEE: No. 5869719artis Corporation

STREET: 520 White Plains Road, P.O. Box 2005

CITY: Tarrytown

STATE: NY

COUNTRY: USA

ZIP: 10591

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30B

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/846,338

FILING DATE:

CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:

NAME: Meigs, J. Timothy

REGISTRATION NUMBER: 38,241

TELECOMMUNICATION INFORMATION:

TELEPHONE: 919-541-8587

TELEFAX: 919-541-8587

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 1041 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

FEATURE:

NAME/KEY: CDS

LOCATION: 1..1038

IDENTIFICATION METHOD: experimental

OTHER INFORMATION: /product= "biotin synthase"

OTHER INFORMATION: /evidence= EXPERIMENTAL

US-08-846-338-7

Query Match 100.0%; Score 25; DB 2; Length 1041;
Best Local Similarity 100.0%; Pred. No. 9.6e-06;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggatctgctgtttgaagcgcagcag 25
|||||
Db 63 GGATCTGCTGTTTGAAGCGCACAG 87

RESULT 3

US-08-411-768B-1

Sequence 1, Application US/08411768B

Patent No. 60837712

GENERAL INFORMATION:

APPLICANT: Olwen Birch

APPLICANT: Johann Blass

APPLICANT: Martin Fuhrmann

APPLICANT: Nicholas Shaw
TITLE OF INVENTION: Biotechnological Method
TITLE OF INVENTION: of Producing Biotin
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:

ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond

STREET: 30 Rockefeller Plaza

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10112

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Wordperfect

SOFTWARE: Version 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/411,768B

FILING DATE: 31-March-95

CLASSIFICATION: 435

PRIORITY APPLICATION DATA:

APPLICATION NUMBER: CH 3124/92

FILING DATE: 02-OCT-1992

PRIORITY APPLICATION DATA:

APPLICATION NUMBER: CH 2134/93

FILING DATE: 15-JUL-1993

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 5872 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: Escherichia coli

STRAIN: DSM498

IMMEDIATE SOURCE:

CLONE: pBO30A-15/9

FEATURE:

NAME/KEY: CDS

LOCATION: 117..1157

IDENTIFICATION METHOD: experimental

OTHER INFORMATION: /codon_start= 117

OTHER INFORMATION: /product= "Biotin synthase"

OTHER INFORMATION: /evidence= EXPERIMENTAL

OTHER INFORMATION: /gene= "biob"

OTHER INFORMATION: /number= 1

FEATURE:

NAME/KEY: CDS

LOCATION: 2295..3050

OTHER INFORMATION: /codon_start= 2295

OTHER INFORMATION: /function= "Involved in pimeloyl-CoA synthesis"

OTHER INFORMATION: /product= "protein"

OTHER INFORMATION: /gene= "bioc"

OTHER INFORMATION: /number= 3

FEATURE:

NAME/KEY: CDS

LOCATION: 3750..5039

IDENTIFICATION METHOD: experimental

OTHER INFORMATION: /codon_start= 3750

OTHER INFORMATION: /EC_number= 2.6.1.62

OTHER INFORMATION: /product= "DAPA synthase"

OTHER INFORMATION: /evidence= EXPERIMENTAL

OTHER INFORMATION: /gene= "bioA"

OTHER INFORMATION: /number= 5

OTHER INFORMATION: /standard_name=

OTHER INFORMATION: "S-adenosyl-L-methionine:8-amino-7-oxononanoate

OTHER INFORMATION: amiotransf."

FEATURE:

NAME/KEY: CDS


```

? LOCATION: 5098..5574
? IDENTIFICATION METHOD: experimental
? OTHER INFORMATION: /codon_start= 5098
? OTHER INFORMATION: /function= "unknown", involved in biotin synthesis"
? OTHER INFORMATION: /product= "protein"
? OTHER INFORMATION: /evidence= EXPERIMENTAL
? OTHER INFORMATION: /gene= "ORF1"
? OTHER INFORMATION: /number= 6
? FEATURE:
? NAME/KEY: -10_signal
? LOCATION: 45..49
? IDENTIFICATION METHOD: experimental
? OTHER INFORMATION: /evidence= EXPERIMENTAL
? OTHER INFORMATION: /standard_name= "promoter ptac"
? FEATURE:
? NAME/KEY: -35_signal
? LOCATION: 23..28
? OTHER INFORMATION: /standard_name= "promoter ptac"
? FEATURE:
? NAME/KEY: RBS
? LOCATION: 105..119
? IDENTIFICATION METHOD: experimental
? OTHER INFORMATION: /evidence= EXPERIMENTAL
? OTHER INFORMATION: /standard_name= "bioB RBS no. 9"
? FEATURE:
? NAME/KEY: RBS
? LOCATION: 2284..2297
? OTHER INFORMATION: /standard_name= "bioC RBS"
? FEATURE:
? NAME/KEY: RBS
? LOCATION: 3742..3752
? OTHER INFORMATION: /standard_name= "bioA RBS"
? FEATURE:
? NAME/KEY: RBS
? LOCATION: 5088..5100
? OTHER INFORMATION: /standard_name= "ORF1 RBS"
? FEATURE:
? NAME/KEY: terminator
? LOCATION: 5583..5644
? OTHER INFORMATION: /standard_name= "rho-independent
? OTHER INFORMATION: transcriptional terminator"
? FEATURE:
? NAME/KEY: stem_loop
? LOCATION: 5583..5605
? FEATURE:
? NAME/KEY: Promoter
? LOCATION: 1..96
? IDENTIFICATION METHOD: experimental
? OTHER INFORMATION: /function= "promoter ptac"
? OTHER INFORMATION: /evidence= EXPERIMENTAL
? PUBLICATION INFORMATION:
? DOCUMENT NUMBER: WO 87/01391 B1
? FILING DATE: 26-AUG-1986
? PUBLICATION DATE: 07-APR-1993
? US-08-411-768B-1
?
? Query Match 100.0%; Score 25; DB 3; Length 5872;
? Best Local Similarity 100.0%; Pred. No. 1e-05;
? Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
?
? QY 1 gagatcgtctgtttgaagcgacagcag 25
? 111111111111111111111111111111
? DB 179 GGATCTGCTGTTGAAGCGACAGCAG 203
?
? RESULT 4
? US-08-411-768B-6
? Sequence 6, Application US/08411768B
? Patent No. 6083712
? GENERAL INFORMATION:
? APPLICANT: Olwen Birch
? APPLICANT: Johann Brass
?
? APPLICANT: Martin Fuhrmann
? APPLICANT: Nicholas Shaw
? TITLE OF INVENTION: Biotechnological Method
? TITLE OF INVENTION: of Producing Biotin
? NUMBER OF SEQUENCES: 19
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Brunhaugh, Graves, Donohue & Raymond
? STREET: 30 Rockefeller Plaza
? CITY: New York
? STATE: New York
? COUNTRY: USA
? ZIP: 10112
? COMPUTER READABLE FORM:
? MEDIUM TYPE: floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Wordperfect
? SOFTWARE: Version 5.1
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/411,768B
? FILING DATE: 31-March-95
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: CH 3124/92
? FILING DATE: 02-OCT-1992
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: CH 2134/93
? FILING DATE: 15-JUL-1993
? INFORMATION FOR SEQ ID NO: 6:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 5872 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: double
? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
? HYPOTHEetical: NO
? ORIGINAL SOURCE:
? ORGANISM: Escherichia coli
? STRAIN: DSM498
? IMMEDIATE SOURCE:
? CLONE: pB030A15-9
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 1154..2308
? IDENTIFICATION METHOD: experimental
? OTHER INFORMATION: /codon_start= 1154
? OTHER INFORMATION: /EC_number= 2.3.1.47
? OTHER INFORMATION: /product= "KAPA synthase"
? OTHER INFORMATION: /evidence= EXPERIMENTAL
? OTHER INFORMATION: /gene= "bioF"
? OTHER INFORMATION: /number= 2
? OTHER INFORMATION: /standard_name= "8-Amino-7-oxononanoate synthase"
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 3043..3753
? IDENTIFICATION METHOD: experimental
? OTHER INFORMATION: /codon_start= 3043
? OTHER INFORMATION: /EC_number= 6.3.3.3
? OTHER INFORMATION: /product= "DTR synthase"
? OTHER INFORMATION: /evidence= EXPERIMENTAL
? OTHER INFORMATION: /gene= "bioD"
? OTHER INFORMATION: /number= 4
? OTHER INFORMATION: /standard_name= "Dethiobiotin synthase"
? FEATURE:
? NAME/KEY: RBS
? LOCATION: 1141..1156
? OTHER INFORMATION: /standard_name= "bioF RBS"
? FEATURE:
? NAME/KEY: RBS
? LOCATION: 3030..3045
? OTHER INFORMATION: /standard_name= "bioD RBS"
? PUBLICATION INFORMATION:
? DOCUMENT NUMBER: WO 87/01391 B1
```

FILING DATE: 26-AUG-1986
PUBLICATION DATE: 07-APR-1993
US-08-411-768B-6

Query Match 100.0%; Score 25; DB 3; Length 5872;
Best Local Similarity 100.0%; Pred. No. 1e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ggaatcgtgttgaagcgcagcag 25
|||||
Db 179 GGATCTGCTTTGAAGCGCAGCAG 203

RESULT 5

US-08-876-991-1
Sequence 1, Application US/08876991

Patent No. 5925360

GENERAL INFORMATION:

APPLICANT: Gregor Meyers, Tillmann R menapf,

APPLICANT: Heinz-J rgen Thiel

TITLE OF INVENTION: Hog cholera virus vaccine and diagnostic

NUMBER OF SEQUENCES: 13

CORRESPONDENCE ADDRESS:

ADDRESSEE: Organon Teknika Corporation

ADDRESSEE: Biotechnology Research Institute

STREET: 1330-A Piccard Drive

CITY: Rockville

STATE: Maryland

COUNTRY: U.S.A.

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/876,991

FILING DATE: 16-JUN-1997

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/747,577

FILING DATE:

APPLICATION NUMBER: US/08/650,584

FILING DATE:

APPLICATION NUMBER: US/08/469,702

FILING DATE:

APPLICATION NUMBER: US/08/123,596

FILING DATE:

APPLICATION NUMBER: 07/797,554

FILING DATE: 22-NOV-1991

APPLICATION NUMBER: US 07/494,991

FILING DATE: 16-MAR-1990

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: William M. Blackstone

REGISTRATION NUMBER: 29,772

REFERENCE/DOCKET NUMBER:

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 258-5200

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 12284 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: cDNA

ORIGINAL SOURCE:

ORGANISM: Hog cholera virus

STRAIN: Alfort

CELL LINE: PK 15 and 3BA1D

FEATURE:

NAME/KEY: CDS

LOCATION: 364..12060
OTHER INFORMATION: /label= 435_KDA.protein

FEATURE:

NAME/KEY: primer_bind

LOCATION: complement (2587..2619)

OTHER INFORMATION: /label= primer_1

FEATURE:

NAME/KEY: primer_bind

LOCATION: complement (2842..2880)

OTHER INFORMATION: /label= primer_2

FEATURE:

NAME/KEY: variation

LOCATION: replace(127, "c")

FEATURE:

NAME/KEY: variation

LOCATION: replace(1522, "g")

FEATURE:

NAME/KEY: variation

LOCATION: replace(10989, "t")

US-08-876-991-1

Query Match 56.0%; Score 14; DB 2; Length 12284;

Best Local Similarity 100.0%; Pred. No. 15;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 tctgctgttgaag 17
|||||

Db 2253 TCTGCTGTTGAAG 2266

RESULT 6

US-09-059-853-1

Sequence 1, Application US/09059853

Patent No. 5935582

GENERAL INFORMATION:

APPLICANT: Gregor Meyers, Tillmann R menapf,

APPLICANT: Heinz-J rgen Thiel

TITLE OF INVENTION: Hog cholera virus vaccine and diagnostic

NUMBER OF SEQUENCES: 13

CORRESPONDENCE ADDRESS:

ADDRESSEE: Organon Teknika Corporation

ADDRESSEE: Biotechnology Research Institute

STREET: 1330-A Piccard Drive

CITY: Rockville

STATE: Maryland

COUNTRY: U.S.A.

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/059,853

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/797,554

FILING DATE: 22-NOV-1991

APPLICATION NUMBER: US 07/494,991

FILING DATE: 16-MAR-1990

ATTORNEY/AGENT INFORMATION:

NAME: William M. Blackstone

REGISTRATION NUMBER: 29,772

REFERENCE/DOCKET NUMBER:

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 258-5200

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 12284 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

```

;
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Hog cholera virus
; STRAIN: Alfort
; CELL LINE: PK 15 and 38A1D
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 364..12060
; OTHER INFORMATION: /label= 435_KDA_protein
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: complement (2587..2619)
; OTHER INFORMATION: /label= primer_1
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: complement (2842..2880)
; OTHER INFORMATION: /label= primer_2
; FEATURE:
; NAME/KEY: replacement(127, "c")
; LOCATION:
; FEATURE:
; NAME/KEY: variation
; LOCATION: replace(1522, "g")
; FEATURE:
; NAME/KEY: variation
; LOCATION: replace(10989, "t")
; US-09-059-853-1

```

```

Query Match          56.0%; Score 14; DB 2; Length 12284;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 4 tctgctgttgaag 17
    |||
DB 2253 tctgctgttgaag 2266

```

```

RESULT 7
US-08-233-608-8/c
; Sequence 8, Application US/08233608
; Patent No. 5585238
; GENERAL INFORMATION:
; APPLICANT: Ligon, James M
; APPLICANT: Beck, James J
; TITLE OF INVENTION: Detection of Fungal Pathogens Using the
; TITLE OF INVENTION: Polymerase Chain Reaction
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233.608
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Spruill, W. Murray
; REGISTRATION NUMBER: 32,943
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8615
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 8:

```

```

;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 17 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Other nucleic acid
; DESCRIPTION: Oligonucleotide primer JB434
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-233-608-8

```

```

Query Match          52.0%; Score 13; DB 1; Length 17;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 12 ttgaagcgcagca 24
    |||
DB 15 TTGAAGCGCAGCA 3

```

```

RESULT 8
US-08-887-480-8/c
; Sequence 8, Application US/08887480
; Patent No. 5814453
; GENERAL INFORMATION:
; APPLICANT: Beck, James J
; TITLE OF INVENTION: Detection of Fungal Pathogens Using the
; TITLE OF INVENTION: Polymerase Chain Reaction
; NUMBER OF SEQUENCES: 96
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5814453artis Corporation
; STREET: 520 White Plains Road
; CITY: Tarrytown
; STATE: NY
; COUNTRY: USA
; ZIP: 10591
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/887,480
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/722,187
; FILING DATE: 15-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: CGC 1739/PCT/CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8689
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 17 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Other nucleic acid
; DESCRIPTION: Oligonucleotide primer JB434
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-887-480-8

```

```

Query Match          52.0%; Score 13; DB 1; Length 17;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 12 ttgaagcgagca 24
|||||
Db 15 TTGAAGCGAGCA 3

RESULT 9

US-08-722-187-8/c
; Sequence 8, Application US/08722187
; Patent No. 5955274
; GENERAL INFORMATION:
; APPLICANT: Ligon, James M
; APPLICANT: Beck, James J
; TITLE OF INVENTION: Detection of Fungal Pathogens Using the
; TITLE OF INVENTION: Polymerase Chain Reaction
; NUMBER OF SEQUENCES: 86
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/722,187
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/233,608
; FILING DATE: 04-APR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Walsh, Andrea C.
; REGISTRATION NUMBER: 34,988
; REFERENCE/DOCKET NUMBER: CGC 1739
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8666
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ. ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 17 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: Oligonucleotide primer JB434
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; PCT-US95-04712-8

Query Match 52.0%; Score 13; DB 2; Length 17;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 ttgaagcgagca 24
|||||
Db 15 TTGAAGCGAGCA 3

RESULT 10

PCT-US95-04712-8/c
; Sequence 8, Application PC/TUS9504712
; GENERAL INFORMATION:
; APPLICANT: Ligon, James M
; APPLICANT: Beck, James J
; TITLE OF INVENTION: Detection of Fungal Pathogens Using the
; TITLE OF INVENTION: Polymerase Chain Reaction
; NUMBER OF SEQUENCES: 86
; CORRESPONDENCE ADDRESS:

ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/04712
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/233,608
; FILING DATE: 04-APR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Walsh, Andrea C.
; REGISTRATION NUMBER: 34,988
; REFERENCE/DOCKET NUMBER: CGC 1739
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8666
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ. ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 17 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Other nucleic acid
; DESCRIPTION: Oligonucleotide primer JB434
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; PCT-US95-04712-8

Query Match 52.0%; Score 13; DB 5; Length 17;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 ttgaagcgagca 24
|||||
Db 15 TTGAAGCGAGCA 3

RESULT 11

US-08-233-608-9/c
; Sequence 9, Application US/08233608
; Patent No. 5585238
; GENERAL INFORMATION:
; APPLICANT: Ligon, James M
; APPLICANT: Beck, James J
; TITLE OF INVENTION: Detection of Fungal Pathogens Using the
; TITLE OF INVENTION: Polymerase Chain Reaction
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,608
; FILING DATE:
; CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: Spruill, W. Murray
REGISTRATION NUMBER: 32,943
REFERENCE/DOCKET NUMBER: CGC 1739
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8615
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 23 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: Oligonucleotide primer JB525
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-233-608-9

Query Match
Best Local Similarity 52.0%; Score 13; DB 1; Length 23;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 ttgaagcgacga 24
|||||
Db 21 TTGAAGCGCAGCA 9

RESULT 12
US-08-887-480-9/c
Sequence 9, Application US/08887480
Patent No. 5814453
GENERAL INFORMATION:
APPLICANT: Beck, James J
TITLE OF INVENTION: Detection of Fungal Pathogens Using the
NUMBER OF SEQUENCES: 96
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5814453artis Corporation
STREET: 520 White Plains Road
CITY: Tarrytown
STATE: NY
COUNTRY: USA
ZIP: 10591
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/887,480
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/722,187
FILING DATE: 15-OCT-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: CGC 1739/PCT/CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 23 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: Oligonucleotide primer JB525
HYPOTHETICAL: NO

ANTI-SENSE: NO
US-08-887-480-9

Query Match
Best Local Similarity 52.0%; Score 13; DB 1; Length 23;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 ttgaagcgacga 24
|||||
Db 21 TTGAAGCGCAGCA 9

RESULT 13
US-08-722-187-9/c
Sequence 9, Application US/08722187
Patent No. 5955274
GENERAL INFORMATION:
APPLICANT: Ligon, James M
APPLICANT: Beck, James J
TITLE OF INVENTION: Detection of Fungal Pathogens Using the
NUMBER OF SEQUENCES: 86
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ciba-Geigy Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/722,187
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/233,608
FILING DATE: 04-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: Walsh, Andrea C.
REGISTRATION NUMBER: 34,988
REFERENCE/DOCKET NUMBER: CGC 1739
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8689
TELEFAX: 919-541-8686
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 23 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: Oligonucleotide primer JB525
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-722-187-9

Query Match
Best Local Similarity 52.0%; Score 13; DB 2; Length 23;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 ttgaagcgacga 24
|||||
Db 21 TTGAAGCGCAGCA 9

RESULT 14
PCT-US95-04712-9/c

```
Sequence 9, Application PC/TUS9504712
GENERAL INFORMATION:
APPLICANT: Ligon, James M
APPLICANT: Beck, James J
TITLE OF INVENTION: Detection of Fungal Pathogens using the
TITLE OF INVENTION: Polymerase Chain Reaction
NUMBER OF SEQUENCES: 86
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ciba-Geigy Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/04712
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/233,608
FILING DATE: 04-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: Walsh, Andrea C.
REGISTRATION NUMBER: 34,968
REFERENCE/DOCKET NUMBER: CGC 1739
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8666
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 23 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: oligonucleotide primer JB525
HYPOTHETICAL: NO
ANTI-SENSE: NO
PCT-US95-04712-9

Query Match          52.0%; Score 13; DB 5; Length 23;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 ttgaagcgacga 24
    |||||||||||
DB 21 TTGAAGCGCAGCA 9

RESULT 15
US-08-150-331-31/c
Sequence 31, Application US/08150331
Patent No. 5516512
GENERAL INFORMATION:
APPLICANT: DORSSERS J. LAMBERTUS C.
APPLICANT: VAN LEEN, ROBERT W.
TITLE OF INVENTION: MUTANS OF HUMAN INTERLEUKIN-3
NUMBER OF SEQUENCES: 48
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 Page Mill Road
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
```

```
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/150,331
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/651,437
FILING DATE: 05-FEB-1991
ATTORNEY/AGENT INFORMATION:
NAME: GRACEY, NANCY J.
REGISTRATION NUMBER: 28,216
REFERENCE/DOCKET NUMBER: 24615-20010.20
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 813-5600
TELEFAX: (415) 494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 27 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-150-331-31
```

```
Query Match          52.0%; Score 13; DB 1; Length 27;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 ttgaagcgacga 24
    |||||||||||
DB 20 TTGAAGCGCAGCA 8
```

```
RESULT 16
US-08-820-170A-8/c
Sequence 8, Application US/08820170A
Patent No. 5831058
GENERAL INFORMATION:
APPLICANT: Tsutomu, FUJIWARA
APPLICANT: Takeshi, MATANABE
APPLICANT: Masato, HORIE
APPLICANT: Toyomasa, KATAGIRI
TITLE OF INVENTION: HUMAN GENE
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas
STREET: 2100 Pennsylvania Avenue, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: United States
ZIP: 20037-3202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/820,170A
FILING DATE:
CLASSIFICATION: 536
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 293-7060
TELEFAX: (202) 293-7860
TELEX: 6491103
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 384 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
```

MOLECULE TYPE: DNA(genomic)
US-08-820-170A-8

Query Match 52.0%; Score 13; DB 2; Length 384;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggaatcgtcgttt 13
|||||
DB 349 GGATCTGCTGTTT 337

RESULT 17
US-09-055-699-8/c
Sequence 8, Application US/09055699
Patent No. 6005088

GENERAL INFORMATION:
APPLICANT: Tsutomu, FUJIWARA
APPLICANT: Takeshi, WATANABE
APPLICANT: Masato, HORIE
APPLICANT: Toyomasa, KATAGIRI
TITLE OF INVENTION: HUMAN GENE
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas
STREET: 2100 Pennsylvania Avenue, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: United States
ZIP: 20037-3202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/055,699
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/820,170
FILING DATE:
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 293-7060
TELEFAX: (202) 293-7860
TELEX: 6491103
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 384 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA(genomic)
US-09-055-699-8

Query Match 52.0%; Score 13; DB 3; Length 384;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggaatcgtcgttt 13
|||||
DB 349 GGATCTGCTGTTT 337

RESULT 18
US-09-273-565-8/c
Sequence 8, Application US/09273565A
Patent No. 6166190
GENERAL INFORMATION:
APPLICANT: FUJIWARA, TSUTOMU
APPLICANT: WATANABE, TAKESHI

APPLICANT: HORIE, MASATO
TITLE OF INVENTION: AN ISOLATED NUCLEIC ACID MOLECULE ENCODING HUMAN
TITLE OF INVENTION: SKELETAL MUSCLE-SPECIFIC UBIQUITIN-CONJUGATING ENZYME
FILE REFERENCE: 0-53599
CURRENT APPLICATION NUMBER: US/09/273,565A
CURRENT FILING DATE: 1999-03-22
EARLIER APPLICATION NUMBER: 09/055,699
EARLIER FILING DATE: 1998-04-07
EARLIER APPLICATION NUMBER: 08/820,170
EARLIER FILING DATE: 1997-03-19
EARLIER APPLICATION NUMBER: JP 63410/1996
EARLIER FILING DATE: 1996-03-19
EARLIER APPLICATION NUMBER: JP 69163/1997
EARLIER FILING DATE: 1997-03-05
NUMBER OF SEQ ID NOS: 95
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 8
LENGTH: 384
TYPE: DNA
ORGANISM: Homo sapiens
US-09-273-565-8

Query Match 52.0%; Score 13; DB 4; Length 384;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggaatcgtcgttt 13
|||||
DB 349 GGATCTGCTGTTT 337

RESULT 19
US-08-233-608-2
Sequence 2, Application US/08233608
Patent No. 5585238
GENERAL INFORMATION:
APPLICANT: Liqon, James M
APPLICANT: Beck, James J
TITLE OF INVENTION: Detection of Fungal Pathogens Using the
NUMBER OF SEQUENCES: 49
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ciba-Geigy Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/233,608
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Spruill, W. Murray
REGISTRATION NUMBER: 32,943
REFERENCE/DOCKET NUMBER: CGC 1739
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8615
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 583 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO

```

; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Septoria nodorum
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..583
; OTHER INFORMATION: /note= "DNA sequence for the
; OTHER INFORMATION: Internal transcribed Spacer of Septoria nodorum"
US-08-233-608-2

Query Match          52.0%; Score 13; DB 1; Length 583;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 ttgaagcgacga 24
Db 464 TTGAAGCGACGA 476

RESULT 20
US-08-887-480-2
; Sequence 2, Application US/08887480
; Patent No. 5814453
; GENERAL INFORMATION:
; APPLICANT: Beck, James J
; TITLE OF INVENTION: Detection of Fungal Pathogens Using the
; TITLE OF INVENTION: Polymerase Chain Reaction
; NUMBER OF SEQUENCES: 96
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: No. 5814453artis Corporation
; STREET: 520 White Plains Road
; CITY: Tarrytown
; STATE: NY
; COUNTRY: USA
; ZIP: 10591
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/887,480
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/722,187
; FILING DATE: 15-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: CGC 1739/PCT/CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8687
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 583 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Septoria nodorum
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..30
; OTHER INFORMATION: /note= "3' end of small subunit
; OTHER INFORMATION: rRNA gene"
; FEATURE:
; NAME/KEY: misc_feature
; ORIGINAL SOURCE:
```

```

; LOCATION: 31..216
; OTHER INFORMATION: /note= "ITS 1"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 217..372
; OTHER INFORMATION: /note= "5.8S rRNA gene"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 373..526
; OTHER INFORMATION: /note= "ITS 2"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 527..583
; OTHER INFORMATION: /note= "5' end of large subunit
; OTHER INFORMATION: rRNA gene"
US-08-887-480-2

Query Match          52.0%; Score 13; DB 1; Length 583;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 ttgaagcgacga 24
Db 464 TTGAAGCGACGA 476

RESULT 21
US-08-722-187-2
; Sequence 2, Application US/08722187
; Patent No. 5955274
; GENERAL INFORMATION:
; APPLICANT: Iigon, James M
; APPLICANT: Beck, James J
; TITLE OF INVENTION: Detection of Fungal Pathogens Using the
; TITLE OF INVENTION: Polymerase Chain Reaction
; NUMBER OF SEQUENCES: 86
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/722,187
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/233,608
; FILING DATE: 04-APR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Walsh, Andrea C.
; REGISTRATION NUMBER: 34,988
; REFERENCE/DOCKET NUMBER: CGC 1739
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8666
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 583 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
```


ORGANISM: Septoria nodorum
FEATURE:
NAME/KEY: misc.feature
LOCATION: 1..583
OTHER INFORMATION: /note="DNA sequence for the
OTHER INFORMATION: Internal Transcribed Spacer of Septoria nodorum"
US-08-722-187-2

Query Match 52.0%; Score 13; DB 2; Length 583;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 ttgaagcgcagca 24
|||||
Db 464 TTGAAGCCGACGCA 476

RESULT 22

PCT-US95-04712-2

Sequence 2, Application PC/TUS9504712

GENERAL INFORMATION:

APPLICANT: Ligon, James M

APPLICANT: Beck, James J

TITLE OF INVENTION: Detection of Fungal Pathogens Using the

TITLE OF INVENTION: Polymerase Chain Reaction

NUMBER OF SEQUENCES: 86

CORRESPONDENCE ADDRESS:

ADDRESS: Ciba-Geigy Corporation

STREET: 7 Skyline Drive

CITY: Hawthorne

STATE: NY

COUNTRY: USA

ZIP: 10532

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/04712

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/233,608

FILING DATE: 04-APR-1994

ATTORNEY/AGENT INFORMATION:

NAME: Walsh, Andrea C.

REGISTRATION/DOCKET NUMBER: 34,988

REFERENCE/DOCKET NUMBER: CGC 1739

TELEPHONE: 919-541-8666

TELEFAX: 919-541-8689

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 583 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

MOLECULE TYPE: DNA (genomic)

HYPOHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE: Septoria nodorum

ORGANISM: Septoria nodorum

FEATURE:

NAME/KEY: misc.feature

LOCATION: 1..583

OTHER INFORMATION: /note="DNA sequence for the

OTHER INFORMATION: Internal Transcribed Spacer of Septoria nodorum"

PCT-US95-04712-2

Query Match 52.0%; Score 13; DB 5; Length 583;

Best Local Similarity 100.0%; Pred. No. 50;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 12 ttgaagcgcagca 24
|||||
Db 464 TTGAAGCCGACGCA 476

RESULT 23

US-08-998-416-727

Sequence 727, Application US/08998416

Patent No. 6239264

GENERAL INFORMATION:

APPLICANT: Philippsen, Peter

APPLICANT: Pohlmann, Rainer

APPLICANT: Steiner, Sabine

APPLICANT: Mohr, Christine

APPLICANT: Wendland, Jürgen

APPLICANT: Knechtel, Philipp

APPLICANT: Reibisch, Corinne

TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPPI

TITLE OF INVENTION: AND USES THEREOF

NUMBER OF SEQUENCES: 1152

CORRESPONDENCE ADDRESS:

ADDRESSEE: NO. 6239264artis Corporation

STREET: 3054 Cornwallis Road

CITY: Research Triangle Park

STATE: No. 6239264th Carolina

COUNTRY: USA

ZIP: 27709

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/998,416

FILING DATE: 24-DEC-1997

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: CH 0016/97

FILING DATE: 31-DEC-1996

ATTORNEY/AGENT INFORMATION:

NAME: Meigs, J. Timothy

REGISTRATION/DOCKET NUMBER: 38,241

REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976

TELEPHONE: 919-541-8689

TELEFAX: 919-541-8687

INFORMATION FOR SEQ ID NO: 727:

SEQUENCE CHARACTERISTICS:

LENGTH: 749 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

MOLECULE TYPE: DNA (genomic)

ORIGINAL SOURCE:

ORGANISM: PAG1475RP

US-08-998-416-727

Query Match 52.0%; Score 13; DB 4; Length 749;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 tgcgttttgaagc 18
|||||
Db 96 TGCTGTTTGAAGC 108

RESULT 24

US-08-153-848-31

Sequence 31, Application US/08153848

Patent No. 5759804
GENERAL INFORMATION:
APPLICANT: Godiska, Ronald
APPLICANT: Gray, Patrick W.
APPLICANT: Schweikart, Vicki L.
TITLE OF INVENTION: No. 5759804e1 Seven Transmembrane Receptors
NUMBER OF SEQUENCES: 64
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/153,848
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/977,452
FILING DATE: 17-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: No. 5759804and, Greta E.
REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 31794
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 474-6300
TELEFAX: (312) 474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 1161 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: exon
LOCATION: 7..80
NAME/KEY: CDS
LOCATION: 94..1158
US-08-153-848-31

Query Match 52.0%; Score 13; DB 1; Length 1161;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 gatctgctgttg 14
|||||

Db 322 GATCTGCTGTTTG 334

RESULT 25
PCT-US93-11153-31
GENERAL INFORMATION:
SEQUENCE 31, Application PC/TUS9311153
APPLICANT: Godiska, Ronald
APPLICANT: Gray, Patrick W.
APPLICANT: Schweikart, Vicki L.
TITLE OF INVENTION: Novel Seven Transmembrane Receptors
NUMBER OF SEQUENCES: 64
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
STREET: 6300 Sears Tower, 233 South Wacker Drive

CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/11153
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/977,452
FILING DATE: 17-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: Noland, Greta E.
REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 31794
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 474-6300
TELEFAX: (312) 474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 1161 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: exon
LOCATION: 7..80
NAME/KEY: CDS
LOCATION: 94..1158
PCT-US93-11153-31

Query Match 52.0%; Score 13; DB 5; Length 1161;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 gatctgctgttg 14
|||||

Db 322 GATCTGCTGTTTG 334

RESULT 26
US-08-252-966B-13/C
SEQUENCE 13, Application US/08252966B
Patent No. 5624818
GENERAL INFORMATION:
APPLICANT: Eisenman, Robert N.
APPLICANT: Hurlin, Peter J.
APPLICANT: Ayer, Donald E.
TITLE OF INVENTION: Regulatory Proteins that Dimerize with
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Christensen, O'Connor, Johnson, and KindnessP LLC
STREET: 1420 Fifth Ave., Suite 2800
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98101-2347
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/252,966B
FILING DATE: 01-JUN-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Shelton, Dennis K.
REGISTRATION NUMBER: 26,997.
REFERENCE/DOCKET NUMBER: FHC17694
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 682-8100
TELEFAX: (206) 224-0779
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 1242 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
DESCRIPTION: clone 10; see Figure 24
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Mus musculus
US-08-252-966B-13

Query Match 52.0%; Score 13; DB 1; Length 1242;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 ctgctgttggaag 17
|||||
DB 525 ctgctgttggaag 513

RESULT 27
US-08-353-550-7/c
Sequence 7, Application US/08353550
PATENT INFORMATION:
PATENT NO. 5744313
APPLICANT: Williams, Lewis T.
APPLICANT: Cavanaugh, William M.
TITLE OF INVENTION: No. 5744313el Protein Domain Which Binds
NUMBER OF SEQUENCES: 9
TITLE OF INVENTION: Tyrosine Phosphorylated Proteins
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend Hourie and Crew
STREET: One Market Plaza, Stewart Tower, Suite 2000
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/353,550
FILING DATE: 09-DEC-1994
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Dow, Karen B.
REGISTRATION NUMBER: 29,684
REFERENCE/DOCKET NUMBER: 02307K-057300
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-326-2400
TELEFAX: 415-326-2422
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1425 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)

FEATURE:
NAME/KEY: CDS
LOCATION: 1..1422
US-08-353-550-7

Query Match 52.0%; Score 13; DB 1; Length 1425;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 tgttgaagcgcga 21
|||||
DB 611 tgttgaagcgcga 599

RESULT 28
US-08-551-687-7/c
Sequence 7, Application US/08551687
PATENT NO. 5925547
GENERAL INFORMATION:
APPLICANT: Williams, Lewis T.
APPLICANT: Cavanaugh, William M.
TITLE OF INVENTION: No. 5925547el Protein Domain Which Binds
NUMBER OF SEQUENCES: 9
TITLE OF INVENTION: Tyrosine Phosphorylated Proteins
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew
STREET: One Market Plaza, Stewart Tower, Suite 2000
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/551,687
FILING DATE: 01-NOV-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/353,550
FILING DATE: 09-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Murphy, Matthew B.
REGISTRATION NUMBER: P39,787
REFERENCE/DOCKET NUMBER: 2307K-5731
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-326-2400
TELEFAX: 415-326-2422
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1425 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1422
US-08-551-687-7

Query Match 52.0%; Score 13; DB 2; Length 1425;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 tgttgaagcgcga 21
|||||
DB 611 tgttgaagcgcga 599

RESULT 29
US-08-820-170A-9/C
; Sequence 9, Application US/08820170A
; Patent No. 5831058
; GENERAL INFORMATION:
; APPLICANT: Tsutomu, FUJIWARA
; APPLICANT: Takeshi, MATANABE
; APPLICANT: Masato, HORIE
; APPLICANT: Toyomasa, KATAGIRI
; TITLE OF INVENTION: HUMAN GENE
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seass
; STREET: 2100 Pennsylvania Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States
; ZIP: 20037-3202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/820,170A
; FILING DATE:
; CLASSIFICATION: 536
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 293-7060
; TELEFAX: (202) 293-7860
; TELEX: 6491103
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1493 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA(genomic)
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; IMMEDIATE SOURCE:
; LIBRARY: Human fetal brain cDNA library
; CLONE: GEN-025F07
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 95..478
; US-08-820-170A-9

Query Match 52.0%; Score 13; DB 2; Length 1493;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggaatcgcgtttt 13
|||||

Db 443 GGATCTGCTGTTT 431

RESULT 30
US-09-055-699-9/C
; Sequence 9, Application US/09055699
; Patent No. 6005088
; GENERAL INFORMATION:
; APPLICANT: Tsutomu, FUJIWARA
; APPLICANT: Takeshi, MATANABE
; APPLICANT: Masato, HORIE
; APPLICANT: Toyomasa, KATAGIRI
; TITLE OF INVENTION: HUMAN GENE
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seass
; STREET: 2100 Pennsylvania Avenue, N.W.
; CITY: Washington

STATE: D.C.
COUNTRY: United States
ZIP: 20037-3202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/055,699
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/820,170
FILING DATE:
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 293-7060
TELEFAX: (202) 293-7860
TELEX: 6491103
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1493 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA(genomic)
HYPOTHEICAL: NO
ANTI-SENSE: NO
IMMEDIATE SOURCE:
LIBRARY: Human fetal brain cDNA library
CLONE: GEN-025F07
FEATURE:
NAME/KEY: CDS
LOCATION: 95..478
US-09-055-699-9

Query Match 52.0%; Score 13; DB 3; Length 1493;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggaatcgcgtttt 13
|||||

Db 443 GGATCTGCTGTTT 431

RESULT 31
US-09-273-565-9/C
; Sequence 9, Application US/09273565A
; Patent No. 6166190
; GENERAL INFORMATION:
; APPLICANT: FUJIWARA, TSUTOMU
; APPLICANT: WATANABE, TAKESHI
; APPLICANT: HORIE, MASATO
; TITLE OF INVENTION: AN ISOLATED NUCLEIC ACID MOLECULE ENCODING HUMAN
; FILE REFERENCE: O-53599
; CURRENT APPLICATION NUMBER: US/09/273,565A
; EARLIER FILING DATE: 1999-03-22
; EARLIER APPLICATION NUMBER: 09/055,699
; EARLIER FILING DATE: 1998-04-07
; EARLIER APPLICATION NUMBER: 08/820,170
; EARLIER FILING DATE: 1997-03-19
; EARLIER APPLICATION NUMBER: JP 63410/1996
; EARLIER FILING DATE: 1996-03-19
; EARLIER APPLICATION NUMBER: JP 69163/1997
; EARLIER FILING DATE: 1997-03-05
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
; LENGTH: 1493
; TYPE: DNA
; ORGANISM: Homo sapiens

FEATURE:
NAME/KEY: CDS
LOCATION: (95)..(478)
US-09-273-565-9

Query Match 52.0%; Score 13; DB 4; Length 1493;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ggaatgctgtt 13
|||||
Db 443 GCATCTGCTGTT 431

RESULT 32
US-08-171-299B-1
Sequence 1, Application US/08171299B
Patent No. 5599665
GENERAL INFORMATION:
APPLICANT: Barbieri, Joseph T.
APPLICANT: Frank, Data W.
TITLE OF INVENTION: EXOENZYME S PROTEIN PREPARATION AND CLONED
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: QUARLES & BRADY
STREET: 411 East Wisconsin Avenue
CITY: Milwaukee
STATE: Wisconsin
COUNTRY: U.S.A.
ZIP: 53202-4497
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/171.299B
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Baker, Jean C.
REGISTRATION NUMBER: 35,433
REFERENCE/DOCKET NUMBER: 650053.90871
TELECOMMUNICATION INFORMATION:
TELEPHONE: (414) 271-5709
TELEFAX: (414) 271-3552
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1712 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-171-299B-1

Query Match 52.0%; Score 13; DB 1; Length 1712;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 tgaagcgacgacg 25
|||||
Db 388 TGAAGCGACGACG 400

RESULT 33
US-09-221-235-9/c
Sequence 9, Application US/09221235
Patent No. 6043040
GENERAL INFORMATION:
APPLICANT: Acton, Susan
TITLE OF INVENTION: NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR

FILE REFERENCE: MNI-050
CURRENT APPLICATION NUMBER: US/09/221.235
CURRENT FILING DATE: 1998-12-28
EARLIER APPLICATION NUMBER: 09/163,115
EARLIER FILING DATE:
NUMBER OF SEQ ID NOS: 15
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 9
LENGTH: 1743
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(1743)
US-09-221-235-9

Query Match 52.0%; Score 13; DB 3; Length 1743;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 tgaagcgacgacg 25
|||||
Db 64 TGAAGCGACGACG 52

RESULT 34
US-09-221-928-9/c
Sequence 9, Application US/09221928
Patent No. 6121030
GENERAL INFORMATION:
APPLICANT: Acton, Susan
TITLE OF INVENTION: NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR
FILE REFERENCE: MNI-050
CURRENT APPLICATION NUMBER: US/09/221.928
CURRENT FILING DATE: 1998-12-28
EARLIER APPLICATION NUMBER: 09/163,115
EARLIER FILING DATE:
NUMBER OF SEQ ID NOS: 15
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 9
LENGTH: 1743
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(1743)
US-09-221-928-9

Query Match 52.0%; Score 13; DB 3; Length 1743;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 tgaagcgacgacg 25
|||||
Db 64 TGAAGCGACGACG 52

RESULT 35
US-09-221-527-9/c
Sequence 9, Application US/09221527
Patent No. 6146832
GENERAL INFORMATION:
APPLICANT: Acton, Susan
TITLE OF INVENTION: NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR
FILE REFERENCE: MNI-050
CURRENT APPLICATION NUMBER: US/09/221.527
CURRENT FILING DATE: 1998-12-28
EARLIER APPLICATION NUMBER: 09/163,115
EARLIER FILING DATE:
NUMBER OF SEQ ID NOS: 15
SOFTWARE: Patentln Ver. 2.0

SEQ ID NO 9
LENGTH: 1743
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(1743)
US-09-221-527-9

Query Match
Best Local Similarity 52.0%; Score 13; DB 3; Length 1743;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 tgaagcgagcag 25
|||||
DB 64 TGAAGCGAGCAG 52

RESULT 36
US-09-221-236-9/c
Sequence 9, Application US/09221236
Patent No. 6146841
GENERAL INFORMATION:
APPLICANT: Acton, Susan
TITLE OF INVENTION: NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR
FILE REFERENCE: MNI-050
CURRENT APPLICATION NUMBER: US/09/221,236
CURRENT FILING DATE: 1998-12-28
EARLIER APPLICATION NUMBER: 09/163,115
EARLIER FILING DATE: 1998-09-29
NUMBER OF SEQ ID NOS: 15
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 9
LENGTH: 1743
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(1743)
US-09-221-236-9

Query Match
Best Local Similarity 52.0%; Score 13; DB 3; Length 1743;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 tgaagcgagcag 25
|||||
DB 64 TGAAGCGAGCAG 52

RESULT 37
US-09-221-416-9/c
Sequence 9, Application US/09221416
Patent No. 6153417
GENERAL INFORMATION:
APPLICANT: Acton, Susan
TITLE OF INVENTION: NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR
FILE REFERENCE: MNI-050
CURRENT APPLICATION NUMBER: US/09/221,416
CURRENT FILING DATE: 1998-12-28
EARLIER APPLICATION NUMBER: 09/163,115
EARLIER FILING DATE: 1998-09-29
NUMBER OF SEQ ID NOS: 15
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 9
LENGTH: 1743
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(1743)

US-09-221-416-9

Query Match
Best Local Similarity 52.0%; Score 13; DB 3; Length 1743;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 tgaagcgagcag 25
|||||
DB 64 TGAAGCGAGCAG 52

RESULT 38
US-09-221-245-9/c
Sequence 9, Application US/09221245
Patent No. 6180358
GENERAL INFORMATION:
APPLICANT: Acton, Susan
TITLE OF INVENTION: NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR
FILE REFERENCE: MNI-050
CURRENT APPLICATION NUMBER: US/09/221,245
CURRENT FILING DATE: 1998-12-28
EARLIER APPLICATION NUMBER: US 09/163,115
EARLIER FILING DATE: 1998-09-29
NUMBER OF SEQ ID NOS: 15
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 9
LENGTH: 1743
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(1743)
US-09-221-245-9

Query Match
Best Local Similarity 52.0%; Score 13; DB 4; Length 1743;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 tgaagcgagcag 25
|||||
DB 64 TGAAGCGAGCAG 52

RESULT 39
US-09-163-115-9/c
Sequence 9, Application US/09163115A
Patent No. 6183962
GENERAL INFORMATION:
APPLICANT: Acton, Susan
TITLE OF INVENTION: NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR
FILE REFERENCE: MNI-050
CURRENT APPLICATION NUMBER: US/09/163,115A
CURRENT FILING DATE: 1998-09-29
NUMBER OF SEQ ID NOS: 15
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 9
LENGTH: 1743
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(1743)
US-09-163-115-9

Query Match
Best Local Similarity 52.0%; Score 13; DB 4; Length 1743;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 tgaagcgagcag 25
|||||

Db 64 TGAAGCGCAGCAG 52

RESULT 40

US-09-221-528-9/c
; Sequence 9, Application US/09221528

; Patent No. 6190874

; GENERAL INFORMATION:

; APPLICANT: Acton, Susan

; TITLE OF INVENTION: NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR

; FILE REFERENCE: MNT-050

; CURRENT APPLICATION NUMBER: US/09/221,528

; CURRENT FILING DATE: 1998-12-28

; EARLIER APPLICATION NUMBER: 09/163,115

; EARLIER FILING DATE: 1998-09-29

; NUMBER OF SEQ ID NOS: 15

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 9

; LENGTH: 1743

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (1)..(1743)

US-09-221-528-9

Query Match 52.0%; Score 13; DB 4; Length 1743;
Best Local Similarity 100.0%; Pred. No. 52;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 tgaagcgacagcag 25

Db 64 TGAAGCGCAGCAG 52

RESULT 41

US-09-593-553-9/c

; Sequence 9, Application US/09593553

; Patent No. 6200770

; GENERAL INFORMATION:

; APPLICANT: Acton, Susan

; TITLE OF INVENTION: NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR

; FILE REFERENCE: MNT-050

; CURRENT APPLICATION NUMBER: US/09/593,553

; CURRENT FILING DATE: 2000-06-14

; PRIOR APPLICATION NUMBER: 09/163,115

; PRIOR FILING DATE: 1998-09-28

; NUMBER OF SEQ ID NOS: 15

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 9

; LENGTH: 1743

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (1)..(1743)

US-09-593-553-9

Query Match 52.0%; Score 13; DB 4; Length 1743;
Best Local Similarity 100.0%; Pred. No. 52;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 tgaagcgacagcag 25

Db 64 TGAAGCGCAGCAG 52

RESULT 42

US-09-221-237-9/c

; Sequence 9, Application US/09221237

; Patent No. 6214597

; GENERAL INFORMATION:

; APPLICANT: Acton, Susan

; TITLE OF INVENTION: NOVEL CSAPK-1 NUCLEIC ACID MOLECULES AND USES THEREFOR

; FILE REFERENCE: MNT-050

; CURRENT APPLICATION NUMBER: US/09/221,237

; CURRENT FILING DATE: 1998-12-28

; EARLIER APPLICATION NUMBER: 09/163,115

; EARLIER FILING DATE: 1998-09-29

; NUMBER OF SEQ ID NOS: 15

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 9

; LENGTH: 1743

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (1)..(1743)

US-09-221-237-9

Query Match 52.0%; Score 13; DB 4; Length 1743;
Best Local Similarity 100.0%; Pred. No. 52;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 tgaagcgacagcag 25

Db 64 TGAAGCGCAGCAG 52

RESULT 43

US-08-361-611-3

; Sequence 3, Application US/08361611

; Patent No. 5519125

; GENERAL INFORMATION:

; APPLICANT: Potter, Sharon L

; APPLICANT: Ward, Eric R

; TITLE OF INVENTION: Plant Adenylosuccinate Synthetase and

; TITLE OF INVENTION: DNA Coding Therefor

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: CIBA-GEIGY Corporation

; STREET: 7 Skyline Drive

; CITY: Hawthorne

; STATE: NY

; COUNTRY: USA

; ZIP: 10532

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30B

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/361,611

; FILING DATE:

; CLASSIFICATION: 536

; ATTORNEY/AGENT INFORMATION:

; NAME: Elmer, James Scott

; REGISTRATION NUMBER: 36,129

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (919) 541-8614

; TELEFAX: (919) 541-8689

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1835 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

; HYPOTHETICAL: NO

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 18..1469

; OTHER INFORMATION: /product= "Maize Adenylosuccinate

OTHER INFORMATION: Synthetase"
US-08-361-611-3

Query Match 52.0%; Score 13; DB 1; Length 1835;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 atctgctgttga 15
|||||
DB 508 ATCTGCTGTTGA 520

RESULT 44

US-08-565-655-3
Sequence 3, Application US/08565655
Patent No. 5688939

GENERAL INFORMATION:

APPLICANT: Potter, Sharon L

APPLICANT: Ward, Eric R

TITLE OF INVENTION: Plant Adenylosuccinate Synthetase and

TITLE OF INVENTION: DNA Coding Therefor

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESS: Ciba Patent Department

STREET: 540 White Plains Rd., POB 2005

CITY: Tarrytown

STATE: NY

COUNTRY: USA

ZIP: 10591-9005

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30B

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/565,655

FILING DATE:

CLASSIFICATION: 210

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/361,611

FILING DATE: 12-DEC-1994

ATTORNEY/AGENT INFORMATION:

NAME: Elmer, James Scott

REGISTRATION NUMBER: 36,129

TELECOMMUNICATION INFORMATION:

TELEPHONE: (919) 541-8614

TELEFAX: (919) 541-8689

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 1835 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

HYPOTHETICAL: NO

FEATURE:

NAME/KEY: CDS

LOCATION: 18..1469

OTHER INFORMATION: /product="Maize Adenylosuccinate

OTHER INFORMATION: Synthetase"

US-08-565-655-3

Query Match 52.0%; Score 13; DB 1; Length 1835;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 atctgctgttga 15
|||||

DB 508 ATCTGCTGTTGA 520

RESULT 45

US-08-946-967-3

Sequence 3, Application US/08946967

Patent No. 5882869

GENERAL INFORMATION:

APPLICANT: Potter, Sharon L

APPLICANT: Ward, Eric R

TITLE OF INVENTION: Plant Adenylosuccinate Synthetase and

TITLE OF INVENTION: DNA Coding Therefor

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESS: CIBA-GEIGY Corporation

STREET: 7 Skyline Drive

CITY: Hawthorne

STATE: NY

COUNTRY: USA

ZIP: 10532

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30B

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/946,967

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Elmer, James Scott

REGISTRATION NUMBER: 36,129

TELECOMMUNICATION INFORMATION:

TELEPHONE: (919) 541-8614

TELEFAX: (919) 541-8689

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 1835 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

HYPOTHETICAL: NO

FEATURE:

NAME/KEY: CDS

LOCATION: 18..1469

OTHER INFORMATION: /product="Maize Adenylosuccinate

OTHER INFORMATION: Synthetase"

US-08-946-967-3

Query Match 52.0%; Score 13; DB 2; Length 1835;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 atctgctgttga 15
|||||

DB 508 ATCTGCTGTTGA 520

Search completed: October 9, 2001, 15:55:27
Job time: 13203 sec

Wed Oct 10 07:45:50 2001

us-09-396-196f-4.oli.rni

Page 19

.

.


```

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*

```

```

190: gb_est110:*
191: gb_est111:*
192: gb_hic:*
193: em_gss_fun:*
194: em_gss_hum1:*
195: em_gss_hum2:*
196: em_gss_hum3:*
197: em_gss_hum4:*
198: em_gss_hum5:*
199: em_gss_hum6:*
200: em_gss_hum7:*
201: em_gss_hum8:*
202: em_gss_hum9:*
203: em_gss_inv1:*
204: em_gss_inv2:*
205: em_gss_inv3:*
206: em_gss_inv4:*
207: em_gss_inv5:*
208: em_gss_inv6:*
209: em_gss_inv7:*
210: em_gss_inv8:*
211: em_gss_inv9:*
212: em_gss_inv10:*
213: em_gss_inv11:*
214: em_gss_inv12:*
215: em_gss_inv13:*
216: em_gss_inv14:*
217: em_gss_inv15:*
218: em_gss_inv16:*
219: em_gss_inv17:*
220: em_gss_inv18:*
221: em_gss_inv19:*
222: em_gss_inv20:*
223: em_gss_inv21:*
224: em_gss_inv22:*
225: em_gss_inv23:*
226: em_gss_inv24:*
227: em_gss_inv25:*
228: em_gss_inv26:*
229: em_gss_inv27:*
230: em_gss_inv28:*
231: em_gss_inv29:*
232: em_gss_inv30:*
233: em_gss_inv31:*
234: em_gss_inv32:*
235: em_gss_inv33:*
236: em_gss_inv34:*
237: em_gss_inv35:*
238: em_gss_inv36:*
239: em_gss_inv37:*
240: em_gss_inv38:*
241: em_gss_inv39:*
242: em_gss_inv40:*
243: em_gss_inv41:*
244: em_gss_inv42:*
245: em_gss_inv43:*
246: em_gss_inv44:*
247: em_gss_inv45:*
248: em_gss_inv46:*
249: em_gss_inv47:*
250: em_gss_inv48:*
251: em_gss_inv49:*
252: em_gss_inv50:*
253: em_gss_inv51:*
254: em_gss_inv52:*
255: em_gss_inv53:*
256: em_gss_inv54:*
257: em_gss_inv55:*
258: em_gss_inv56:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	17	68.0	479	136	BE508108	BE508108 dc26f09.y
2	16	64.0	293	240	AZ214122	AZ214122 Sheared D
3	16	64.0	358	17	A1207085	A1207085 qf59a09.x
4	16	64.0	360	156	C61070	C61070 C61070 Yuj1
5	16	64.0	502	136	BE515828	BE515828 WHE0606.B
6	16	64.0	517	232	A0692054	A0692054 HS-5408_A
7	16	64.0	555	235	A0889100	A0889100 HS-2204_A
8	16	64.0	558	111	AW064984	AW064984 ST38C02.P
9	16	64.0	567	165	BE232412	BE232412 137424.MA
10	16	64.0	632	244	AZ449344	AZ449344 1M0247J04
11	16	64.0	650	258	TA69C05P	TA69C05P T. brucei
12	17	66.0	222	177	FR0044714	FR0044714 Fugu rubr
13	15	60.0	278	15	A1021028	A1021028 ua99b01.r
14	15	60.0	298	131	BB337949	BB337949 BB337949
15	15	60.0	316	222	FR0039786	FR0039786 Fugu rubr
16	15	60.0	358	147	BF355075	BF355075 RC2-HT082
17	15	60.0	372	229	A0521765	A0521765 HS-5239.A
18	15	60.0	406	224	A0143500	A0143500 HS-3066_B
19	15	60.0	424	229	A0494880	A0494880 HS-5195_A
20	15	60.0	435	224	A0114851	A0114851 C1T-HSP-2
21	15	60.0	437	116	AW432846	AW432846 sh81905.Y
22	15	60.0	456	150	BF555582	BF555582 UI-R-E0-C
23	15	60.0	466	20	A1477772	A1477772 fbs4h11.Y
24	15	60.0	495	225	A0192174	A0192174 HS-3227_B
25	15	60.0	507	163	BE114310	BE114310 UI-R-E01-
26	15	60.0	509	122	AW915326	AW915326 EST34630
27	15	60.0	528	122	AA818607	AA818607 UI-R-A0-b
28	15	60.0	532	258	TA375R11P	TA375R11P T. brucei
29	15	60.0	542	12	AA790463	AA790463 wv04g12.r
30	15	60.0	559	258	TA162C10P	TA162C10P T. brucei
31	15	60.0	594	222	FR0036852	FR0036852 Fugu rubr
32	15	60.0	598	222	FR0036852	FR0036852 Fugu rubr
33	15	60.0	600	222	FR0044579	FR0044579 Fugu rubr
34	15	60.0	611	222	FR0044579	FR0044579 Fugu rubr
35	15	60.0	621	175	BC306774	BC306774 fm0608.Y
36	15	60.0	627	172	BC017095	BC017095 df68903.x
37	15	60.0	632	233	AO754510	AO754510 HS-5434_A
38	15	60.0	640	155	BC546580	BC546580 60257435
39	15	60.0	645	137	BE597936	BE597936 PT1-67_C0
40	15	60.0	669	173	BC085090	BC085090 H3108E05-
41	15	60.0	663	118	AW633493	AW633493 b108d10.W
42	15	60.0	681	140	BE820949	BE820949 GM700013A
43	15	60.0	748	112	AM128741	AM128741 fe36h03.Y
44	15	60.0	752	220	CNS02060	AM128741 Tetraodon
45	15	60.0	846	153	BC428569	BC428569 602501093

ALIGNMENTS

RESULT 1
BE508108/c
LOCUS
DEFINITION
sequence.
BE508108 479 bp mRNA
dc26f09.y1 NICHHD XGC L11 Xenopus laevis cDNA clone IMAGE:3398249 5'
similar to TR:075940 075940 30KDA SPLICING FACTOR.; mRNA

ACCESSION BE508108
VERSION BE508108
KEYWORDS EST.
SOURCE African clawed frog.
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae;
Xenopodinae; Xenopus.
1 (bases 1 to 479)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
REFERENCE
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE

JOURNAL
COMMENT
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapsb-remail.nih.gov

Tumor Gene Index
Tissue Procurement: Martha Rebert, Steven L. Klein, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: Xenopus clones from this library are available
through the I.M.A.G.E. Consortium/LLNL at: info@image.llnl.gov
Seq primer: -40RP from Gibco
High quality sequence stop: 434.
Location/Qualifiers

FEATURES

source
1. 479
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone_image="3398249"
/clone_lib="NICHHD XGC L11"
/lab_host="DH10B (phage-resistant)"
/note="Organ: liver; Vector: pCMW-SPORT6; site_1: NotI;
site_2: SalI; cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.4 kb. Constructed by Life
Technologies. Note: This is a Xenopus Gene Collection (XGC
) library."
BASE COUNT 114 a 147 c 112 g 106 t
ORIGIN

Query Match 68.0%; Score 17; DB 136; Length 479;
Best Local Similarity 100.0%; Pred. No. 7.5;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 tggttgaagcgagcagcagc 25
|||||
DB 183 TGGTTGAAGCGAGCAGCAGC 167

RESULT 2
LOCUS
DEFINITION
Sheared DNA-67G1.TF Sheared DNA Trypanosoma brucei genomic clone
A2214122
A2214122 GSS 09-JUN-2000

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Trypanosoma brucei.
Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma.

REFERENCE
AUTHORS
El-sayed, N., Zhao, S., Zhao, H., Gill, S., Suh, E., Malek, J., Fujii, C.,
Gerrard, C., Leech, V., de Jong, P., Ullu, E., Melville, S., Donelson, J.,
Fraser, C. and Adams, M.
Determination of clone end sequences from Trypanosoma brucei GUTat
10.1 sheared DNA library
Unpublished (1999)
Other GSSs: Sheared DNA-67G1.TF
Contact: Najib M. El-sayed
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208

EMAIL: nelsayed@tigr.org
Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared
DNA library constructed at TIGR. Clones will be available for
distribution through Research Genetics, Alabama, USA. Sheared DNA
end sequences search page: http://www.tigr.org/cdb/mbd/tbdt/
Seg primer: M13-Forward
Class: Shotgun.

FEATURES

source
location/Qualifiers
1. 293
/organism="Trypanosoma brucei"

/strain="TRE0927/4 GUTat 10.1"
 /db_xref="taxon:5691"
 /clone="Sheared DNA-67G1"
 /clone_id="Sheared DNA"
 /note="Vector: pUC18; Site 1: SmaI; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from a cloned population of Trypanosoma brucei (TRE0927/4 GUTat 10.1) was mechanically sheared to give a tight size distribution (approx 2 kb). The v + 1 method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds M. Vaubin and B. Barrell, Oxford University Press, 1999)."

BASE COUNT 38 a 73 c 78 g 104 t
 ORIGIN

Query Match 64.0%; Score 16; DB 240; Length 293;
 Best Local Similarity 100.0%; Pred. No. 28;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 atctgctgttgaagc 18
 |||||
 Db 149 ATCTGCTGTTGAAGC 164

RESULT 3
 LOCUS AI207085 358 bp mRNA EST 10-NOV-1998
 DEFINITION qf59a09.x1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1754296
 3' similar to TR:061655 Q61655 EUKARYOTIC TRANSLATION INITIATION
 FACTOR 4A RELATED SEQUENCE 1; mRNA sequence.

ACCESSION AI207085
 VERSION AI207085
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 358)
 NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 CONTACT: Robert Strausberg, Ph.D.
 Email: cgaaps-r@mail.nih.gov
 CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo
 , Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CCGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 www.bio.llnl.gov/bbrp/image/image.html
 Insert Length: 652 Std Error: 0.00
 Seq primer: -400P from glbco
 High quality sequence stop: 239.
 Location/Qualifiers
 1..358

FEATURES
 source

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1754296"
 /clone_id="Soares_testis_NHT"
 /sex="male"
 /lab_host="DH10B"
 /note="Vector: pT7m3D-Pac (Pharmacia) with a modified
 polylinker. Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
 was prepared from mRNA obtained from Clontech Laboratories
 , Inc., and primed with a Not I - Oligo(dT) primer [5'
 TGTTACCAATCTGAGTGGAGCGCCCAATTTTCTTTTCTTTT 3']
 Double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I

and Eco RI sites of the modified pT7m3 vector. Library
 went through one round of normalization to Cot5, and was
 constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 98 a 79 c 73 g 107 t
 ORIGIN

Query Match 64.0%; Score 16; DB 17; Length 358;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 ctgttgaagcagc 23
 |||||
 Db 294 CTGTTGAAGCGCAGC 309

RESULT 4
 LOCUS C61070/c 360 bp mRNA EST 22-SEP-1997
 DEFINITION C61070 Yui Kohara unpublished cDNA Caenorhabditis elegans cDNA
 clone YK224c7 5', mRNA sequence.
 ACCESSION C61070
 VERSION C61070.1 GI:2419775
 KEYWORDS EST.
 SOURCE Caenorhabditis elegans.
 ORGANISM Caenorhabditis elegans.
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea
 ; Rhabditidae; Pelodierinae; Caenorhabditis.
 1 (bases 1 to 360)
 Kohara Y., Motohashi, T., Tabara, H., Watanabe, H., Sugimoto, A., Sano
 , M., Miyata, A. and Nishigaki, A.
 Expression map of the C.elegans genome
 Unpublished (1996)
 CONTACT: Yui Kohara
 Gene Library Lab
 National Institute of Genetics
 Yata 1111, Mishima, Shizuoka 411, Japan
 Tel: 81-559-81-6854
 Fax: 81-559-81-6855
 Email: ykohara@lab.nig.ac.jp.

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

FEATURES
 source
 1..360
 /organism="Caenorhabditis elegans"
 /strain="CB1409 him 6(e1489)"
 /db_xref="taxon:6239"
 /clone="YK224c7"
 /clone_id="Yui Kohara unpublished cDNA"
 /sex="hermaphrodite, male"
 /tissue.type="whole animal"
 /dev-stage="varied"

BASE COUNT 101 a 71 c 76 g 111 t 1 others
 ORIGIN

Query Match 64.0%; Score 16; DB 156; Length 360;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 atctgctgttgaagc 18
 |||||
 Db 26 ATCTGCTGTTGAAGC 11

RESULT 5
 LOCUS BE515828 502 bp mRNA EST 08-AUG-2000
 DEFINITION WHE0606_B07_C14Za wheat ABA-treated embryo cDNA library Triticum
 aestivum cDNA clone WHE0606_B07_C14, mRNA sequence.
 ACCESSION BE515828
 VERSION BE515828.1 GI:9739842
 KEYWORDS EST.
 SOURCE bread wheat.
 ORGANISM Triticum aestivum

JOURNAL
MEDLINE
COMMENT

Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@jeng.med.bufileo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.bufileo.edu/ordering_bac.htm) or from Resear. h Genetics (info@resgen.com). BAC end Web Server: <http://www.htsc.washington.edu>
Plate: 964 Row: 1 Column: 14
Seq primer: T7
Class: BAC ends
High quality sequence stop: 517.

FEATURES
Source
Location/Qualifiers
1..502
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=984 Col=14 Row=1"
/clone_1lb="RPCI-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBAC3.6 vector at EcoRI sites"

BASE COUNT
154 a 96 c 83 g 178 t 6 others

ORIGIN

Query Match 64.0%; Score 16; DB 232; Length 517;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 atctgctgttgagc 18
|||||

Db 411 ATCTGCTGTTGAACG 426

RESULT 7
LOCUS A0889100 555 bp DNA GSS 10-NOV-1999
DEFINITION HS_2204.AL.C07.MR.C17 Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=2204 Col=13 Row=E, DNA sequence.
ACCESSION A0889100
VERSION A0889100.1 GI:6345290
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 555)
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome

JOURNAL
MEDLINE
COMMENT

Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones may be purchased from Resear. h Genetics (info@resgen.com). BAC end Web Server: <http://www.htsc.washington.edu>

FEATURES
Source
Location/Qualifiers
1..502
/organism="Triticum aestivum"
/cultivar="Brevor (Soft, white, winter, common wheat)"
/db_xref="taxon:4565"
/clone="WHE0606_B07_C14"
/clone_1lb="Wheat ABA-treated embryo cDNA library"
/tissue-type="Seed embryo"
/dev_stage="Mature dormant seeds"
/lab_host="E. coli DH12S"
/note="Vector: pGAD10; Site_1: EcoRI; Site_2: XhoI; Embryos were cut from mature, dormant seeds and imbedded in 25 microm ABA (abscisic acid) in 5 mM Mes buffer, pH 5.7, for 12 hr at 22 C. The tissue, total RNA, and poly(A) RNA were prepared by Steven Verhey in M.K. Walker-Simmons's lab (USDA-ARS, Washington State Univ., Pullman, Washington 99164-6420. A cDNA library was made by Clontech using a combination of random and oligo dr primers. Library was plated and archived by Russell Johnson (Colby College, ME/Walker-Simmons' lab). Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."

BASE COUNT
103 a 108 c 128 g 162 t 1 others

ORIGIN

Query Match 64.0%; Score 16; DB 136; Length 502;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 atctgctgttgagc 18
|||||

Db 189 ATCTGCTGTTGAACG 204

RESULT 6
LOCUS A0692054 517 bp DNA GSS 06-JUL-1999
DEFINITION HS_5408.A2.E07.T7A RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate=984 Col=14 Row=I, DNA sequence.
ACCESSION A0692054
VERSION A0692054.1 GI:5382302
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 517)
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome

Plate: 2204 row: E column: 13
 Seq primer: M13 Reverse
 Class: BAC ends
 High quality sequence stop: 555.
 Location/Qualifiers
 1..555

FEATURES
 source
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="plate-2204 Col-13 Row=E"
 /clone_id="CIT Approved Human Genomic Sperm Library D"
 /sex="male"
 /note="Organ: sperm; Vector: pBelobAC11; BAC Clones in E-Coli DH10B"
 BASE COUNT 131 a 135 c 147 g 124 t 18 others
 ORIGIN

Query Match 64.0%; Score 16; DB 235; Length 555;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 tctctgtttgaacgcg 19
 |||||
 DB 497 TCTGCTTTGAACGC 482

RESULT 8
 AM064984 558 bp mRNA EST 12-OCT-1999
 LOCUS
 DEFINITION ST38C02 Pine triplex shoot tip library Pinus taeda cDNA clone
 ACCESSION AM064984
 VERSION AM064984.1 GI:6020056
 KEYWORDS EST.
 SOURCE 10b101ly pine.
 ORGANISM Pinus taeda
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.
 1 (bases 1 to 558)
 Whetten, R.W., Kinlaw, C.S., Retzel, E. and Sederoff, R.R.
 The Pine Gene Discovery Project
 Unpublished (1999)
 CONTACT: Ross Whetten
 Forest Biotechnology Group
 North Carolina State University
 Dept. of Forestry, NC State University, 6113 Jordan Hall, Raleigh
 NC, 27695-8008
 Tel: 919-515-7800
 Fax: 919-515-7801
 Email: rosswhetten@ncsu.edu
 Seq primer: 5' lambda triplex2 Sequencing Primer.

FEATURES
 source
 Location/Qualifiers
 1..558
 /organism="Pinus taeda"
 /db_xref="taxon:3352"
 /clone="ST38C02"
 /clone_lib="Pine triplex shoot tip library"
 /lab_host="E. coli BM25.8"
 /note="Organ: shoot tips; Vector: lambda triplex; Site_1:
 Site_2: Site_3: Site_4: Site_5: Site_6: Site_7: Site_8: Site_9: Site_10:
 apex) were collected during the spring, frozen and used
 for mRNA isolation. The SMART-PCR method (Clontech) was
 used to prepare a library from 1 ug total RNA, using the
 lambda triplex vector. Plasmid subclones in pTriplex were
 recovered by cre-lox excision in E. coli strain BM25.8 and
 sequenced from the 5' end."

BASE COUNT 146 a 117 c 131 g 160 t 4 others
 ORIGIN

Query Match 64.0%; Score 16; DB 111; Length 558;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 gtttgaacgcgcacg 25
 |||||
 DB 108 GTTTGAAGCGACGACG 123

RESULT 9
 BE232412 567 bp mRNA EST 10-JUL-2000
 LOCUS
 DEFINITION 137424 MARC 1p1g Sus scrofa cDNA 5', mRNA sequence.
 ACCESSION BE232412
 VERSION BE232412.1 GI:9017130
 KEYWORDS EST.
 SOURCE pig.
 ORGANISM Sus scrofa
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 1 (bases 1 to 567)
 Fahrenkrug, S.C., Freking, B.A., Rohrer, G.A., Smith, T.P.L., Casas, E.,
 Stone, R.T., Heaton, M.P., Grosse, W.M., Bennett, G.A., Laegreid, W.W.
 and Keele, J.W.
 Design and use of two pooled tissue normalized cDNA libraries for
 EST discovery in swine
 Unpublished (2000)
 CONTACT: Smith TPL
 USDA, ARS, US Meat Animal Research Center
 PO Box 166, Clay Center, NE 68933-0166, USA
 Tel: 402 762 4366
 Fax: 402 762 4390
 Email: smith@email.marc.usda.gov
 Single pass sequencing. Bases called and alt. trimmed with phred
 v0.980904.e. Vector identified by cross-match with the -m10score 18
 and -mismatch 12 options.
 PCR Primers
 FORWARD: AGGAACAGCTATGACCAT
 BACKWARD: GTTTCACGACGACG
 Plate: 72 row: E column: 24
 Seq primer: ATTGAGTGACATATAG.
 Location/Qualifiers
 1..567
 /organism="Sus scrofa"
 /db_xref="taxon:9823"
 /clone_lib="MARC 1p1g"
 /tissue_type="pooled"
 /lab_host="DH10B"
 /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
 Library made from pooled tissue from day 11, 13, 15, 20,
 and 30 embryos."

FEATURES
 source
 Location/Qualifiers
 1..567
 /organism="Sus scrofa"
 /db_xref="taxon:9823"
 /clone_lib="MARC 1p1g"
 /tissue_type="pooled"
 /lab_host="DH10B"
 /note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
 Library made from pooled tissue from day 11, 13, 15, 20,
 and 30 embryos."

BASE COUNT 114 a 198 c 171 g 83 t 1 others
 ORIGIN

Query Match 64.0%; Score 16; DB 165; Length 567;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 ctgtttgaacgcgcacg 23
 |||||
 DB 227 CTGTTTGAAGCGACGACG 212

RESULT 10
 A2449344 632 bp DNA GSS 04-OCT-2000
 LOCUS
 DEFINITION 1M0247J04R Mouse 10kb plasmid U06C1M library Mus musculus genomic
 clone U06C1M0247J04 R, DNA sequence.
 ACCESSION A2449344
 VERSION A2449344.1 GI:10603038
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

Query Match 64.0%; Score 16; DB 165; Length 567;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

REFERENCE 1 (bases 1 to 632)
 AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duyal,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tinney,A., von Niederhausern,A. and Wright,D., Weiss,R.
 TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
 JOURNAL Unpublished (2000)
 COMMENT Contact: Robert B. Weiss
 University of Utah Genome Center
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: dunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0247 ROW: J Column: 04
 Seq primer: CACACAGAAACAGCTATGACC
 Class: plasmid ends
 High quality sequence stop: 632.
 Location/Qualifiers
 1..632
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UUCGIM0247J04"
 /clone_1lb="Mouse 10kb plasmid UUCGIM library"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: PMD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PMD42 (gII4732114[gb|AF129072.1]) a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into 'chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 185 a 167 c 110 g 170 t
 ORIGIN

Query Match 64.0%; Score 16; DB 244; Length 632;
 Best Local Similarity 100.0%; Pred. No. 30;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggatctgctgttgaac 16
 ||||||||||||||||
 Db 552 GGATCTGCTGTGAA 567

RESULT 11
 TA69C05P 650 bp DNA GSS 13-DEC-2000
 LOCUS T. brucei sheared genomic DNA clone 69c05, forward sequence,
 DEFINITION genomic survey sequence.
 ACCESSION AL457556
 VERSION AL457556.1 GI:11858782
 KEYWORDS GSS.
 SOURCE Trypanosoma brucei.
 ORGANISM Trypanosoma brucei
 Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;

REFERENCE 1 (bases 1 to 650)
 AUTHORS Hall,N., Bowman,S., Lennard,N.J., Doggett,J., Atkin,R., Chillingworth,C., Ormond,D., Harris,B., El-Sayed,N., Hou,L., Melville,S.E., Rajandream,M.A. and Barrell,B.G.
 TITLE Direct Submission
 JOURNAL Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and nh@sanger.ac.uk
 COMMENT Constructed at the Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from a cloned population of Trypanosoma brucei (TREN927/4 G9rat 10.1) was mechanically sheared to give a tight size distribution (to give the library construction is 4 kb). The v + i method used for the library construction is described in detail in Smith, H. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999).
 Email: nelsayed@tigr.org
 Details of T. brucei sequencing at the Sanger Centre are available at http://www.sanger.ac.uk/Projects/T_brucei/.
 Location/Qualifiers
 1..650
 /organism="Trypanosoma brucei"
 /strain="TREN927"
 /db_xref="taxon:5691"
 /clone="69c05"
 /clone_1lb="cosmid 192G14"

BASE COUNT 208 a 186 c 173 g 83 t
 ORIGIN

Query Match 64.0%; Score 16; DB 258; Length 650;
 Best Local Similarity 100.0%; Pred. No. 30;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 atctgctgttgaac 18
 ||||||||||||||||
 Db 357 ATCTGCTGTGAAAC 342

RESULT 12
 FR0044714 177 bp DNA GSS 22-OCT-1999
 LOCUS Fugu rubripes GSS sequence, clone 192G14eb6, genomic survey
 DEFINITION sequence.
 ACCESSION AL132206
 VERSION AL132206.1 GI:6114152
 KEYWORDS GSS; genome survey sequence.
 SOURCE Takifugu rubripes.
 ORGANISM Takifugu rubripes
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei;
 Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
 Tetraodontidae; Takifugu.
 1 (bases 1 to 177)
 Elgar,G., Clark,M.S., Smith,S., Meek,S., Warner,S., Edwards,Y.J.K., Umranta,Y., Williams,G. and Brenner,S.
 Direct Submission
 Submitted (11-OCT-1999) MRC Human Genome Mapping Project Resource Centre, Hinxton, Cambridge, CB10 1SB, UK Email:
 biohelp@hmp.mrc.ac.uk
 Vector: plnuescript II KS
 V type: phagemid
 PRIMER: KS
 DESC: One pass dye-terminator sequencing of cosmid cloned genomic
 sequence.
 Location/Qualifiers
 1..177
 /organism="Takifugu rubripes"
 /db_xref="taxon:31033"
 /clone_1lb="cosmid 192G14"

ACCESSION A0521765.1 GI:4752890
 VERSION A0521765.1
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 COMMENT Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 372)
 AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
 TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
 JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 MEDLINE 99380589
 COMMENT Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Plietier de Jong (pietier@edj.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm) or from Resear h Genetics (info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu
 Plate: 815 row: 0 column: 18
 Seq primer: SP6
 Class: BAC ends
 High quality sequence stop: 372.
 FEATURES
 source
 Location/Qualifiers
 1..372
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="Plate=815 Col=18 Row=0"
 /clone_lib="RPCI-11 Human Male BAC Library"
 /sex="male"
 /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at EcoRI sites"
 BASE COUNT 97 a 86 c 72 g 115 t 2 others
 ORIGIN
 Query Match 60.0%; Score 15; DB 229; Length 372;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 2 gatctctgtttgaa 16
 Db 315 GATCTGCTGTTTCAA 329
 RESULT 18
 A0143500 406 bp DNA GSS 24-SEP-1998
 LOCUS HS_3066_B2_H10_MR_C1T Approved Human Genomic Sperm Library D Homo
 DEFINITION sapiens genomic clone Plate=3066 Col=20 Row=P, DNA sequence.
 ACCESSION A0143500
 VERSION A0143500.1 GI:3534153
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 COMMENT Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 406)
 AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
 JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 MEDLINE 99380589
 COMMENT Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Plietier de Jong (pietier@edj.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm) or from Resear h Genetics (info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu
 Plate: 771 row: C column: 11
 Seq primer: SP6
 Class: BAC ends
 FEATURES
 source
 Location/Qualifiers
 1..406
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="Plate=3066 Col=20 Row=P"
 /clone_lib="C1T Approved Human Genomic Sperm Library D"
 /sex="male"
 /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in E-Coli DH10B"
 BASE COUNT 138 a 78 c 54 g 135 t 1 others
 ORIGIN
 Query Match 60.0%; Score 15; DB 224; Length 406;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 2 gatctctgtttgaa 16
 Db 98 GATCTGCTGTTTCAA 84
 RESULT 19
 A0494880 424 bp DNA GSS 28-APR-1999
 LOCUS HS_5195_A1_B06_SP6E RPCI-11 Human Male BAC Library Homo sapiens
 DEFINITION genomic clone Plate=771 Col=11 Row=C, DNA sequence.
 ACCESSION A0494880
 VERSION A0494880.1 GI:4695047
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 COMMENT Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 424)
 AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.
 TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
 JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 MEDLINE 99380589
 COMMENT Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Plietier de Jong (pietier@edj.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm) or from Resear h Genetics (info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu
 Plate: 771 row: C column: 11
 Seq primer: SP6
 Class: BAC ends

RESULT	23
AI47772/c	
LOCUS	AI47772
DEFINITION	466 bp mRNA
	EST
	09-MAR-1999
	fb5h11.y1 zebrafish WASHU MPIMG EST Danio rerio cdna 5', mRNA sequence.

```

97 a 135 c 120 g 113 t 1 others
BASE COUNT
ORIGIN
/organism="Danio rerio"
/db_xref="taxon:7955"
/clone_lib="Zebrafish WashU MPMG E57"
/sex="mixed"
/tissue_type="26 somite embryos, adult livers, shield
stage embryos"
/lab_host="Xil-blue MRF"
/notes="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; 1st
strand cDNA was primed with a Not I - oligo(dt)15 primer
[5' pGACATGATCTGTAGATCGGACGCGCCGCTTTTTTTTTT3']
double-stranded cDNA was ligated to Sal I adaptors (BRL),
digested with Not I and cloned into the Not I and Sal I
sites of the pSPORT1 vector (BRL). Library was constructed
by Matthew Clark (Lehrnach lab, ICRF, London and Max Planck
Institut fuer Molekulare Genetik, Berlin). cDNAs for E57
analysis were selected following oligonucleotide
hybridization fingerprinting of arrayed clones from
zebrafish late somitogenesis (26 ss), adult liver or
embryonic shield stage (5,6 h) libraries. Fingerprint
data were used to computationally cluster cDNAs, and a
single cDNA from each cluster was chosen for sequencing.
In some cases multiple members of the same cluster were
sequenced to assess clustering parameters or single clones
were sequenced additional times to assess quality
control."
```

	60.0%;	Score 15;	DB 20;	Length 466;
Query Match				
Best Local Similarity	100.0%;	Pred. No. 1.le+2;		
Matches 15; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;
QY 8 cgttttgaagcgacg 22				
Dd 330 ctcttttgaaaccgcag 316				

RESULT 24
 A0192174 495 bp DNA GSS 04-NOV-1998
 LOCUS HS.3227.B1.G01.T7 CIT Approved Human Genomic Sperm Library D Homo
 DEFINITION sapiens genomic clone Plate=3227 Col=1 Row=N, DNA sequence.
 ACCESSION A0192174
 VERSION A0192174.1 GI:3590796
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 495)
 Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
 Kellar,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
 Hood,L.
 Sequence-tagged connectors: A sequence approach to mapping and
 scanning the human genome
 Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 JOURNAL 99380589
 MEDLINE Contact: Mahairas GG, Wallace JC, Hood L
 COMMENT High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Sequence tagged connector
 Plate: 3227 Row: N column: 1
 Class: BAC ends
 High quality sequence stop: 495.
 Location/Qualifiers
 1. 495
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="Plate=3227 Col=1 Row=N"
 /clone_lib="CIT Approved Human Genomic Sperm Library D"
 /sex="male"
 /note="Organ: sperm; Vector: pBelobAC11; BAC Clones in
 E-Coli DH10B"
 BASE COUNT 135 a 114 c 79 g 167 t
 ORIGIN
 Query Match 60.0%; Score 15; DB 225; Length 495;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 2 gatctgctgtttgaa 16
 |||||||||||||
 Db 316 GATCTGCTGTGAA 330
 RESULT 25
 BE114310 507 bp mRNA EST 13-JUN-2000
 LOCUS BE114310/c
 DEFINITION UI-R-BJ1-awk-g-02-0-UI.s1 UI-R-BJ1 Rattus norvegicus cDNA clone
 UI-R-BJ1-awk-g-02-0-UI 3', mRNA sequence.
 ACCESSION BE114310
 VERSION BE114310.1 GI:8506415
 KEYWORDS EST.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 507)
 Bonaldo,M.F., Lennon,G. and Soares,M.B.
 Normalization and subtraction: two approaches to facilitate gene
 discovery
 Genome Res. 6 (9), 791-806 (1996)
 JOURNAL 97044477
 MEDLINE Contact: Soares, MB
 COMMENT

Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msoares@blue.weeg.uiowa.edu
 The sequence contained an oligo-dT track that was present in the
 oligonucleotide that was used to prime the synthesis of first
 strand cDNA and therefore this may represent a bonafide poly A
 tail. The sequence tag present in the cDNA between the NotI site
 and the oligo-dT track served to identify it as a clone from the
 normalized heart library cDNA library preparation: M.B. Soares Lab
 Clone distribution: clones will be available through Research
 Genetics (www.resgen.com)
 Seq primer: M13 forward
 POLYA=Yes.
 FEATURES
 source location/Qualifiers
 1. 507
 /organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone="UI-R-BJ1-awk-g-02-0-UI"
 /clone_lib="UI-R-BJ1"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: p7T3D-Pac (Pharmacia) with a modified
 polylinker. Site_1: Not I; Site_2: Eco RI; The UI-R-BJ1
 library is a subtracted library derived from the following
 tissues: atrium at 16.5 dpc, ventricle at 16.5 dpc, AV
 canal at 16.5 dpc, ventricle at 13 dpc, and adult heart.
 For a detailed description of the library from which this
 clone was derived, please visit our web site at
 ratseq.eng.uiowa.edu. The subtraction has been previously
 described in (Bonaldo, Lennon and Soares, Genome Research
 6:791-806, 1996)
 TAG_LIB=UI-R-BJ1
 TAG_TISSUE=heart
 TAG_SEQ=ACAC"
 BASE COUNT 125 a 119 c 105 g 158 t
 ORIGIN
 Query Match 60.0%; Score 15; DB 163; Length 507;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 4 tctgctgtttgaac 18
 |||||||||||||
 Db 357 TCTGCTGTGAAAC 343
 RESULT 26
 AW915326 509 bp mRNA EST 25-MAY-2000
 LOCUS AW915326/c
 DEFINITION EST34630 Normalized rat embryo. Bento Soares Rattus sp. cDNA clone
 RGT0011 5' end, mRNA sequence.
 ACCESSION AW915326
 VERSION AW915326.1 GI:8081017
 KEYWORDS EST.
 SOURCE Rattus sp.
 ORGANISM Rattus sp.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 509)
 Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
 Kerlavage,A.R. and Adams,M.D.
 Rat Genome Project: Generation of a Rat EST (RESE) Catalog & Rat
 Gene Index
 Unpublished (1998)
 JOURNAL
 COMMENT The Institute for Genomic Research
 9712, Medical Center Drive, Rockville, MD 20850, USA

Tel: (301)-838-3529
Fax: (301)-838-0208

Email: nhlee@igf.org

This clone is available through the ATCC, contact the ATCC
tel:703-365-2700 for further information

Seq primer: M13 Reverse.

FEATURES

source

Location/Qualifiers
1. 509
/organism="Rattus sp."
/db_xref="taxon:10118"
/clone="RGIC011"
/clone_id="Normalized rat embryo", Bento Soares"
/dev_stage="embryo 8, 12, 18 dpc"
/note="Vector: pT73Pac; Site_1: EcoRI; Site_2: NotI"

BASE COUNT

116 a 167 c 118 g 107 t 1 others

ORIGIN

Query Match
Best Local Similarity 100.0%; Score 15; DB 122; Length 509;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggaatcgtgttga 15
|||||
Db 443 GGATCTGCTGTTGA 429

RESULT 27

AA818607/c

LOCUS

DEFINITION

AA818607 528 bp mRNA EST 03-JUL-1999
UI-R-A0-bc-h-05-0-UI s1 UI-R-A0 Rattus norvegicus cDNA clone
UI-R-A0-bc-h-05-0-UI 3' similar to gb|U96683|MMU96683 Mus musculus
immunoglobulin-like receptor PIRAZ (6M23) mRNA, partial cds, mRNA
sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
AA818607.1 GI:4228398
EST.
Norway rat.
Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
COMMENT
1 (bases 1 to 528)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
On Feb 17, 1998 this sequence version replaced gi:2889346.
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9665
Email: msosares@blue.weeg.uiowa.edu

FEATURES

source

Location/Qualifiers
1. 528
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-A0-bc-h-05-0-UI"
/clone_id="UI-R-A0"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"

/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; This library
consists of a mixture of individually tagged normalized
libraries constructed from rat placenta, adult lung, brain
liver, kidney, heart, spleen, ovary, and muscle. The tag
is a string of 3-5 nucleotides present between the Not I
site and the oligo-dT track which allows identification of
the library of origin of a clone within the mixture."

BASE COUNT

138 a 131 c 101 g 158 t

Query Match
Best Local Similarity 100.0%; Score 15; DB 12; Length 528;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 tctgctgttgaagc 18
|||||
Db 353 TCTGCTGTTGAAGC 339

RESULT 28

TA375E11P/c

LOCUS

DEFINITION

TA375E11P 532 bp DNA GSS 13-DEC-2000
T. brucei sheared genomic DNA clone 375e11, forward sequence,
genomic survey sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
AL495591.1 GI:11873033
GSS.
Trypanosoma brucei.
Trypanosoma brucei.
Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
1 (bases 1 to 532)
Hall,N., Bowman,S., Leonard,N.J., Doggett,J., Atkin,R.,
Chillingworth,C., Ormond,D., Harris,B., El-Sayed,N., Hou,L.,
Melville,S.E., Rajadream,M.A. and Barrell,B.G.
Direct Submission
Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
project, Sanger Centre, The Wellcome Trust Genome Campus,
Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
nhlee@sanger.ac.uk

Constructed at the Institute for Genomic Research (TIGR),
Rockville, MD. Genomic DNA isolated from a cloned population of
Trypanosoma brucei (TRU927/4 GUTat 10.1) was mechanically sheared
to give a tight size distribution (4 kb). The v + i method used for the library construction is
described in detail in Smith, H. and Venter, J.C. (Making small
insert libraries for whole genome shotgun sequencing projects. In
Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barrell, Oxford University Press, 1999).
Email: nhlee@sanger.ac.uk
Details of T. brucei sequencing at the Sanger Centre are available
at http://www.sanger.ac.uk/Projects/T_brucei/.

FEATURES

source

Location/Qualifiers
1. 532
/organism="Trypanosoma brucei"
/strain="TRU927"
/db_xref="taxon:5691"
/clone="375e11"

BASE COUNT
162 a 156 c 92 g 122 t

ORIGIN

Query Match
Best Local Similarity 100.0%; Score 15; DB 258; Length 532;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 ctgctgttgaagcg 19
|||||
Db 310 CTGCTGTTGAAGCG 296

RESULT 29
AA790463/c 542 bp mRNA EST 06-FEB-1998
LOCUS
DEFINITION AA790463
IMAGE:1230982.5, similar to SW:RAP3_HUMAN P17964 RAS-RELATED
PROTEIN RAP-2B. ; mRNA sequence.

ACCESSION
VERSION AA790463.1 GI:2850583
SOURCE
ORGANISM Mus musculus
KEYWORDS house mouse.
REFERENCE
AUTHORS Mammalia: Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 542)
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Stepcoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
The WashU-HMI Mouse EST Project
Unpublished (1996)
CONTACT: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.wustl.edu
This clone is available royalty-free through LBNL; contact the
IMAGE Consortium (info@image.lbnl.gov) for further information.
MG1:656574
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 492.

FEATURES
source
1. 542
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:1230982"
/clone_id="Soares_mammary_gland_NBMGC"
/sex="male"
/tissue_type="mammary gland"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Organ: mammary gland. Vector: pT773D-Pac (Pharmacia
) with a modified polylinker. Site_1: Not I. Site_2: Eco
RI. 1st strand cDNA was primed with a Not I - oligo(dT)
primer 15'.
TCTTACCAATCTGAGTGGAGCGCGCGAATGTTTTTTTTTTTTTTTTTTTTT
T 3'; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), sites of the modified pT773 vector.
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
constructed and normalized by Bento Soares and M. Fatima
Bonaldo."

BASE COUNT 127 a 178 c 125 g 112 t
ORIGIN

Query Match 60.0%; Score 15; DB 12; Length 542;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggatcgcgtttga 15
|||||
Db 486 ggatcgcgtttga 472

RESULT 30
TA162C10F/c 559 bp DNA GSS 13-DEC-2000
LOCUS
DEFINITION T. brucei sheared genomic DNA clone 162c10, forward sequence,
genomic survey sequence.

ACCESSION
VERSION AL472440.1 GI:11837892
KEYWORDS GSS.
SOURCE Trypanosoma brucei.
ORGANISM Trypanosoma brucei
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;

REFERENCE
AUTHORS 1 (bases 1 to 559)
Hall, N., Bowman, S., Lennard, N.J., Doggett, J., Atkin, R.,
Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L.,
Melville, S.E., Rajandream, M.A. and Barrell, B.G.
Direct Submission
Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
Cambridge CB10 1SA. E-mail: barrell@sanger.ac.uk and
nhlesanger@ac.uk
Constructed at the Institute for Genomic Research (TIGR),
Rockville, MD. Genomic DNA isolated from a cloned population of
Trypanosoma brucei (TREU927/4 GUTAT 10.1) was mechanically sheared
to give a tight size distribution (4 kb). The v + i method used for the library construction is
described in detail in Smith, H. and Venter, J.C. (Making small
insert libraries for whole genome shotgun sequencing projects. In
Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barrell, Oxford University Press, 1999).
Email: neilsayed@tigr.org
Details of T. brucei sequencing at the Sanger Centre are available
at <http://www.sanger.ac.uk/projects/T-brucei/>.

FEATURES
source
1. 559
Location/Qualifiers
/organism="Trypanosoma brucei"
/strain="TREU927"
/db_xref="taxon:5691"
/clone="162c10"

BASE COUNT 198 a 154 c 92 g 115 t
ORIGIN

Query Match 60.0%; Score 15; DB 258; Length 559;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 ctgcctgttgaagcg 19
|||||
Db 188 ctgcctgttgaagcg 174

RESULT 31
FR0036852 594 bp DNA GSS 22-OCT-1999
LOCUS
DEFINITION Fugu rubripes GSS sequence, clone 039N10aE9, genomic survey
sequence.
VERSION AL124355
KEYWORDS GSS; genome survey sequence.
SOURCE
ORGANISM Takifugu rubripes.
Takifugu rubripes.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthopterygii; Acanthopterygii; Percormphae; Tetraodontiformes;
Tetraodontidae; Takifugu.
1 (bases 1 to 594)
Elgar, G., Clark, M.S., Smith, S., Meek, S., Warner, S., Edwards, Y.J.K.,
Umranta, Y., Williams, G. and Brenner, S.
Direct Submission
Submitted (11-OCT-1999) MRC Human Genome Mapping Project Resource
Centre, Hinxton, Cambridge, CB10 1SB. UK Email:
biolhelp@hmp.mrc.ac.uk
Vector: pBluescript II KS
V_type: phagemid
PRIMER: KS
DESCR: One pass dye-terminator sequencing of cosmid cloned genomic

FEATURES sequence.

LOCATION/Qualifiers

1. 594

/organism="Takifugu rubripes"

/db_xref="taxon:31033"

/clone_lib="cosmid 039N10"

/clone="039N10aE9"

BASE COUNT 138 a 135 c 144 g 157 t 20 others

ORIGIN

Query Match 60.0%; Score 15; DB 222; Length 594;

Best Local Similarity 100.0%; Pred. No. 1.1e+02;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 atcgcgtgttgaag 17

Db 475 ATCTGCTGTTGAAG 489

RESULT 32

FR0035666 598 bp DNA GSS 22-OCT-1999

LOCUS Fugu rubripes GSS sequence, clone 019B02CC2, genomic survey

DEFINITION

SEQUENCE

AL123183

AL123183.1 GI:6104798

VERSION

KEYWORDS

SOURCE

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Takifugu.

1 (bases 1 to 598)

Elgar,G., Clark,M.S., Smith,S., Meek,S., Warner,S., Edwards,Y.J.K., Umrana,Y., Williams,G. and Brenner,S.

Direct Submission

Submitted (11-OCT-1999) MRC Human Genome Mapping Project Resource Centre, Hinxton, Cambridge, CB10 1SB. UK Email:

biohelp@hmp.mrc.ac.uk

Vector: pluescript II KS

V_type: phagemid

PRIMER: KS

DESCR:

One pass dye-terminator sequencing of cosmid cloned genomic sequence.

LOCATION/Qualifiers

1. 598

/organism="Takifugu rubripes"

/db_xref="taxon:31033"

/clone_lib="cosmid 019B02"

/clone="019B02CC2"

BASE COUNT 120 a 147 c 153 g 165 t 13 others

ORIGIN

Query Match 60.0%; Score 15; DB 222; Length 598;

Best Local Similarity 100.0%; Pred. No. 1.1e+02;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 atcgcgtgttgaag 17

Db 116 ATCTGCTGTTGAAG 130

RESULT 33

FR0044579 600 bp DNA GSS 22-OCT-1999

LOCUS Fugu rubripes GSS sequence, clone 192G14AF9, genomic survey

DEFINITION

SEQUENCE

AL132071

AL132071.1 GI:6114017

VERSION

KEYWORDS

GSS: genome survey sequence.

Takifugu rubripes.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Takifugu.

1 (bases 1 to 600)

Elgar,G., Clark,M.S., Smith,S., Meek,S., Warner,S., Edwards,Y.J.K., Umrana,Y., Williams,G. and Brenner,S.

Direct Submission

Submitted (11-OCT-1999) MRC Human Genome Mapping Project Resource Centre, Hinxton, Cambridge, CB10 1SB. UK Email:

biohelp@hmp.mrc.ac.uk

Vector: pluescript II KS

V_type: phagemid

PRIMER: KS

DESCR:

One pass dye-terminator sequencing of cosmid cloned genomic sequence.

LOCATION/Qualifiers

1. 600

/organism="Takifugu rubripes"

/db_xref="taxon:31033"

/clone_lib="cosmid 192G14"

/clone="192G14AF9"

BASE COUNT 123 a 152 c 153 g 170 t 2 others

ORIGIN

Query Match 60.0%; Score 15; DB 222; Length 600;

Best Local Similarity 100.0%; Pred. No. 1.1e+02;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 atcgcgtgttgaag 17

Db 145 ATCTGCTGTTGAAG 159

RESULT 34

FR0044632 611 bp DNA GSS 22-OCT-1999

LOCUS Fugu rubripes GSS sequence, clone 192G14BC7, genomic survey

DEFINITION

SEQUENCE

AL132124

AL132124.1 GI:6114070

VERSION

KEYWORDS

SOURCE

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Takifugu.

1 (bases 1 to 611)

Elgar,G., Clark,M.S., Smith,S., Meek,S., Warner,S., Edwards,Y.J.K., Umrana,Y., Williams,G. and Brenner,S.

Direct Submission

Submitted (11-OCT-1999) MRC Human Genome Mapping Project Resource Centre, Hinxton, Cambridge, CB10 1SB. UK Email:

biohelp@hmp.mrc.ac.uk

Vector: pluescript II KS

V_type: phagemid

PRIMER: KS

DESCR:

One pass dye-terminator sequencing of cosmid cloned genomic sequence.

LOCATION/Qualifiers

1. 611

/organism="Takifugu rubripes"

/db_xref="taxon:31033"

/clone_lib="cosmid 192G14"

/clone="192G14BC7"

BASE COUNT 158 a 154 c 150 g 122 t 27 others

ORIGIN

Query Match 60.0%; Score 15; DB 222; Length 611;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 atctgctgttgaag 17
 |||||
 DB 450 ATCTGCTGTGAG 436

RESULT 35
 BG306774 621 bp mRNA EST 22-FEB-2001
 LOCUS f06b08.y1 zebrafish adult retina cDNA clone
 DEFINITION 4145535 5', similar to TR-0920X1 0920X1 APOPTOSIS-INDUCING FACTOR
 ALT.: mRNA sequence.

ACCESSION BG306774 GI:13104301
 VERSION BG306774.1
 KEYWORDS EST.
 SOURCE zebrafish.
 ORGANISM Danio rerio.

REFERENCE Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
 Actinopterygii: Neopterygii: Teleostei: Euteleostei: Ostariophysi:
 Cypriniformes: Cyprinidae: Rasbora: Danio.
 1 (bases 1 to 621)
 Clark, M., Johnson, S.L., Lehrach, H., Lee, R., Li, F., Marra, M., Eddy
 'S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood
 'K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B.,
 Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Riter, E.,
 Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R.
 and Wilson, R.

TITLE Washu zebrafish EST Project 1998
 JOURNAL Unpublished (1998)
 COMMENT Contact: Stephen L. Johnson
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810

Email: zbrfish@wustl.edu
 Library constructed by: Susan E. Brockerhoff DNA Sequencing by:
 Washington University Genome Sequencing Center Clone distribution:
 RessourcenzentrumPrimaatendank, Berlin, Germany (web address:
 www.rzpd.de)
 Seq primer: T3 ET from Amersham
 High quality sequence stop: 420.

FEATURES
 source Location/Qualifiers
 1..621

/organism="Danio rerio"
 /strain="wild-type"
 /db_xref="taxon:7955"
 /clone="4145535"
 /clone_lib="zebrafish adult retina cDNA"
 /sex="mixed"
 /dev_stage="1-2 years"
 /lab_host="E.Coli XL1-Blue MRF' (XL1-Blue MRF')"
 /note="Vector: Lambda ZAP II (pBluescript SK-); Site: 1:
 EcoRI; Site: 2: SalI; This Zebrafish library was
 constructed by Dr. Susan E. Brockerhoff (email:
 sbrocker@u.washington.edu) RZPD library number: 760"
 BASE COUNT 157 a 141 c 164 g 154 t 5 others
 ORIGIN

Query Match 60.0%; Score 15; DB 175; Length 621;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 8 ctgttgaagcag 22
 |||||
 DB 180 CTGTTGAAGCCAG 166

RESULT 36
 BG017095/c 627 bp mRNA EST 24-JAN-2001
 LOCUS df68g03.x1 Xenopus laevis unfertilized egg cDNA library Xenopus
 DEFINITION laevis cDNA clone XENOPUS_SOURCE_ID:xlneeg005n05 3', mRNA sequence.
 ACCESSION BG017095 GI:12470841

VERSION BG017095.1 GI:12470841
 KEYWORDS EST.
 SOURCE African clawed frog.
 ORGANISM Xenopus laevis

REFERENCE Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:
 Amphibia: Batrachia: Anura: Mesobatrachia: Pipidae: Pipidae:
 Xenopodinae: Xenopus.
 1 (bases 1 to 627)
 Clifton, S., Johnson, S.L., Blumberg, B., Song, J., Hillier, L., Pape, D.,
 Clifton, S., Wylie, T., Underwood, K., Theising, B., Bowers, Y., Person
 'B., Gibbons, M., Harvey, N., Riter, E., Jackson, Y., McCann, R.,
 Waterston, R. and Wilson, R.
 Washu Xenopus EST project, 1999
 Unpublished (1999)
 Other ESTs: df68g03.y1
 Contact: Sandy Clifton, Ph.D.
 Washu Xenopus EST project, 1999
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810

Email: est@wustl.edu
 Library constructed by Bruce Blumberg
 Library normalized by Jihwan Song
 DNA sequencing by: Washington University Genome Sequencing Center
 Clone distribution: Xenopus clone distribution information for
 this library can be found through Research Genetics, visit their
 web page at: <http://www.resgen.com/> Please reference the id listed
 below when ordering this clone: Source lab clone id - xlneeg005n05
 Seq primer: 40UP from Glbco
 High quality sequence stop: 468.

FEATURES
 source Location/Qualifiers
 1..627

/organism="Xenopus laevis"
 /db_xref="taxon:8355"
 /clone="XENOPUS_SOURCE_ID:xlneeg005n05"
 /clone_lib="Xenopus laevis unfertilized egg cDNA library"
 /tissue_type="unfertilized egg"
 /lab_host="Top-10 F"
 /note="Vector: pBluescript SK-; Site: 1: EcoRI; Site: 2:
 XhoI; cDNA was prepared from 2ug of poly A+ RNA.
 XhoI: cDNA was then ligated into unlabeled XR
 (Stratagene) with EcoRI at the 5' end and XhoI at the 3'
 end. SS-library phagemids were prepared by mass excision
 from the original library and normalized by hybridization
 to biotinylated driver (prepared from the same library by
 PCR) to Cot-omega of 11. After removal of hybrids and
 excess driver by streptavidin sepharose chromatography,
 the ss-phagemids were made double stranded and
 electroporated into Top-10 F'. Original library
 constructed by Bruce Blumberg (Blumberg et al., 1991
 Science 253, 194-196; Hawley et al., 1995, Genes Dev. 9,
 2923-2935). Note: This is a Xenopus Gene Collection (XGC)
 library."
 BASE COUNT 146 a 131 c 143 g 206 t 1 others
 ORIGIN

Query Match 60.0%; Score 15; DB 172; Length 627;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 11 tttaagcagcag 25
 |||||
 DB 556 TTGAAGCAGCAG 542

RESULT 37
A0754510 632 bp DNA GSS 27-JUL-1999
LOCUS HS_5434.AL.G05.SP66.RPCI-11 Human Male BAC Library Homo sapiens
DEFINITION genomic clone plate=1010 Col-9 Row-M, DNA sequence.
ACCESSION A0754510
VERSION A0754510.1 GI:5618726
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 632)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
or from Research Genetics (<http://www.resgen.com>). BAC end Web Server:
<http://www.resgen.com>
Plate: 1010 row: M column: 9
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 632.
Location/Qualifiers
1..632
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Plate=1010 Col=9 Row=M"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBACE3.6; Site.1: EcoRI; Site.2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI-Methylase. Size selected DNA was cloned into the
pBACE3.6 vector at EcoRI sites"

BASE COUNT 156 a 143 c 126 g 190 t 17 others
ORIGIN

Query Match 60.0%; Score 15; DB 233; Length 632;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 gctgttgagcgca 21
|||||
Db 308 GCTGTTGAAGCGCA 322

RESULT 38
BG546580
LOCUS BG546580 640 bp mRNA EST 04-APR-2001
DEFINITION 602574435f1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4702473 5',
RNA sequence.
ACCESSION BG546580
VERSION BG546580.1 GI:13545245
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens.

REFERENCE 1 (bases 1 to 640)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-rt@mail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLCM1539 row: f column: 10
High quality sequence stop: 640.
Location/Qualifiers
1..640
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:4702473"
/clone_lib="NIH_MGC_77"
/lab_host="DH10B (TI phage-resistant)"
/note="Organ: Lung; Vector: pDNR-LIB (Clontech); Site.1:
SfiI (ggccgctggcc); Site.2: SfiI (ggccatggcc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CAGCGCCATTATGCGC-3' and 3' adaptor
sequence: 5'-ATTCTAGAGCGGAGCGCGCCGACATG-dT(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.9
kb (range 0.5-4.0 kb). 12/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA). Note: this is a NIH MGC Library."

BASE COUNT 170 a 131 c 155 g 184 t
ORIGIN

Query Match 60.0%; Score 15; DB 155; Length 640;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 atctgtgttgag 17
|||||
Db 600 ATCTGCTTTGAAG 614

RESULT 39
BE597936
LOCUS BE597936 645 bp mRNA EST 18-AUG-2000
DEFINITION P11_67_C08_g1.A002 Pathogen induced 1 (P11) Sorghum bicolor cDNA,
RNA sequence.
ACCESSION BE597936
VERSION BE597936.1 GI:9853009
KEYWORDS EST.
SOURCE sorghum.
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 645)
AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Dean,R., Sudman,M. and Pratt
L.H.
TITLE An EST database from Sorghum: pathogen-induced plants
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions

below Phred quality 16. The threshold for highest quality sequence is 20.
 Seq primer: POLYTMix
 High quality sequence start: 3
 High quality sequence stop: 635
 POLYA=No.

FEATURES

source Location/Qualifiers

1.645
 /organism="Sorghum bicolor"
 /db_xref="taxon:4558"
 /clone_lib="Pathogen induced 1 (PI1)"
 /note="Organ: Anthracnose-infected leaves from two-week-old sorghum plants 48 hr after inoculation: vector: phluescript II from lambda zap II; Site_1: XhoI; Site_2: EcoRI; Two-week-old sorghum plants (BTX 623 cultivar) were infected with pathogen (isolate PRM21 of Colletotrichum graminicola, which is a sorghum isolate). RNA was prepared from infected leaves harvested from 45 seedlings (2 weeks old) exhibit juvenile resistant reaction, which is an incompatible interaction. As they grow older (4 weeks or older), plants resume susceptibility to anthracnose disease. The library was made from poly-A RNA in the cloning vector lambda zap II. Clones to be sequenced were prepared by mass excision. WARNING: While most or all ESTs are expected to derive from the host plant, no effort was made to eliminate ESTs deriving from the pathogen."

BASE COUNT 179 a 143 c 170 g 153 t
 ORIGIN

Query Match 60.0%; Score 15; DB 137; Length 645;
 Best Local Similarity 100.0%; Pred. NO. 1.1e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 ctgctgtttgaagc 19
 |||||||||||||
 Db 627 CTGCTGTTGAAGC 641

RESULT 40
 BG085090/c 649 bp mRNA EST 26-JAN-2001
 LOCUS BG085090-5 NIA Mouse 15K cDNA Clone Set Mus musculus cDNA clone
 DEFINITION H3108E05 5', mRNA sequence.
 ACCESSION BG085090
 VERSION BG085090.1 GI:12567654
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 649)
 AUTHORS Kargul, G.J., Dudekula, D.B., Qian, Y., Lim, M.K., Jandarat, S.A., Tanaka, T.S., Carter, M.G. and Ko, M.S.H.
 TITLE Verification and initial annotation of NIA mouse 15K cDNA clone set
 JOURNAL Unpublished (2001)
 COMMENT Other_ESTS: H3108E05-3
 Contact: George J. Kargul
 Laboratory of Genetics
 National Institute on Aging/National Institutes of Health
 333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
 Email: cdnaelg@sun.grc.nia.nih.gov
 This clone set has been freely distributed to the community. Please visit <http://lgsun.grc.nia.nih.gov/cDNA/15k.html> for details.
 Plate: H3108 row: E column: 05
 Seq primer: -21M13 Reverse
 High quality sequence stop: 649
 POLYA=No.

FEATURES
 source Location/Qualifiers
 1.649
 /organism="Mus musculus"

/strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="H3108E05"
 /clone_lib="NIA Mouse 15K cDNA Clone Set"
 /sex="Clones arrayed from a variety of cDNA libraries"
 /dev_stage="Clones arrayed from a variety of cDNA libraries"
 /lab_host="DH10B"

/note="Vector: pSPORT1, Site_1: SalI, Site_2: NotI. This clone is among a rearranged set of 15,247 clones from 11 embryo cDNA libraries (including preimplantation stage embryos from unfertilized egg to blastocyst, embryonic part of E7.5 embryos, extraembryonic part of E7.5 embryos, and E12.5 female mesonephros/gonad) and one newborn ovary cDNA library. Average insert size 1.5 kb. All source cDNA libraries are cloned unidirectionally with Oligo(dT)-Not primers. References include: (1) Genome-wide expression profiling of mid-gestation placenta and embryo using a 15,000 mouse developmental cDNA microarray, 2000, Proc. Natl. Acad. Sci. U.S.A. 97: 9127-9132; (2) Large-scale cDNA analysis reveals phased gene expression patterns during preimplantation mouse development, 2000, Development, 127: 1737-1749; (3) Genome-wide mapping of unselected transcripts from extraembryonic tissue of 7.5-day mouse embryos reveals enrichment in the t-complex and under-representation on the X chromosome, 1998, Hum Mol Genet 7: 1967-1978."

BASE COUNT 157 a 182 c 177 g 133 t
 ORIGIN

Query Match 60.0%; Score 15; DB 173; Length 649;
 Best Local Similarity 100.0%; Pred. NO. 1.1e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ggatcgtcgtttga 15
 |||||||||||||
 Db 315 GGATCGCTGTTGA 301

RESULT 41
 AM633493 663 bp mRNA EST 03-APR-2000
 LOCUS AM633493
 DEFINITION laevis cDNA clone pBX0008D10 5', mRNA sequence.
 ACCESSION AM633493
 VERSION AM633493.1 GI:7390574
 KEYWORDS EST.
 SOURCE African clawed frog.
 ORGANISM Xenopus laevis.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipridae; Xenopodinae; Xenopus.
 REFERENCE 1 (bases 1 to 663)
 AUTHORS Blackhear, P.J., Lai, W.S., Thorn, J.M., Kennington, E.A., Staffa, N.G., Jr., Moore, D.T., Bouffard, G.G., Beckstrom-Sternberg, S.M., Touchman, J.W., Bonaldo, M.F. and Soares, M.B.
 TITLE The NIEHS Xenopus Maternal EST Project
 JOURNAL Unpublished (2000)
 COMMENT Contact: Perry J. Blackhear
 Office of Clinical Research and Laboratory of Signal Transduction
 National Institute of Environmental Health Sciences
 A2-05 NIEHS, 101 Alexander Drive, Research Triangle Park, NC 27709, USA
 Tel: 919 541-4899
 Fax: 919 541-4571
 Email: black009@niehs.nih.gov
 Clone is available through Research Genetics, Inc., 2130 Memorial Parkway, Huntsville, AL 35891
 Phone: 800-533-4363 ext.cdna, fax 256-536-9016 att:cdna, email cdna@resgen.com
 DNA Sequencing and analyses performed by National Institutes of Health Intramural Sequencing Center (NISC).

PCR Primers

FORWARD: TGTAAACGACGCGCAGT
 BACKWARD: CAGGAACAGCTATGACC
 Plate: 0008 row: D column: 10
 Seq primer: T7 primer.

FEATURES

source

Location/Qualifiers
 1. 663
 /organism="Xenopus laevis"
 /db_xref="taxon:8355"
 /clone_lib="PBX0008D10"
 /clone_lib="Blackshear/Soares normalized Xenopus egg library"
 /sex="female"
 /tissue_type="unfertilized egg"
 /cell_type="unfertilized egg"
 /dev_stage="unfertilized egg"
 /lab_host="DH10B"
 /note="Vector: pT73-Pac; Site:1: EcoRI; Site:2: NotI; POLYA-selected mRNA was prepared from unfertilized Xenopus laevis eggs. The library was constructed in the vector pT73-Pac as described in Bonaldo, M.F., Lennon, G. and Soares, M.B. Normalization and subtraction: two approaches to facilitate gene discovery", Genome Research 6:791-806, 1996. The first strand synthesis used a NotI-dri8 primer; double stranded cDNAs were ligated to EcoRI adapters, digested with NotI, and directionally cloned into the NotI and EcoRI-digested pT73-Pac vector. The library contained approximately 7.2 x 10⁵ recombinants, with average insert sizes of 1-1.5 kb."

BASE COUNT
 211 a 172 c 139 g 140 t

ORIGIN
 Query Match 60.0%; Score 15; DB 118; Length 663;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 ttctgaagcgcagcag 25
 Db 237 TTTGAAGCGCAGCAG 251

RESULT 42
 BE820949/c

LOCUS BE820949 681 bp mRNA EST 21-SEP-2000
 DEFINITION GM700013A20C12 Gm-r1070 Glycine max cDNA clone Gm-r1070-5112 3',
 mRNA sequence.
 BE820949
 BE820949.1 GI:10253183

ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

soybean.
 Glycine max
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.
 1 (bases 1 to 681)
 Voklin, L., Keim, P., Shoemaker, R., Retzel, E., Khanna, A., Corvelli, V.,
 Expanding, J., Rapp, C., Shoop, E., Pardinas, J., Liu, L. and Lewin, H.
 A Functional Genomics Program for Soybean (NSF 9872565)
 Unpublished (1999)
 Other ESTs: AM432846 corresponding to Gm-cl016-6609 (5')
 Contact: Voklin, L.O., PI, A Functional Genomics Program for
 Soybean (NSF 9872565)
 Lewin, H. A., Director, Keck Center for Comparative and Functional
 Genomics
 University of Illinois
 Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA
 Tel: (217) 244-6147
 Fax: (217) 333-4582
 Email: l-voklin@uiuc.edu
 This clone is available through: Genome Systems, Inc. 4633 World
 Parkway Circle St. Louis, Missouri 63134. For further information
 call: (800) 430-0030 pr (314) 427-3222 FAX: (888) 919-3524 or (314)

427-3324 or contact: clones@genomesystems.com or info@genome
 systems.com web site: www.genomesystems.com
 Seq primer: 5'-TTTTTTTTTTTTTTTTT(A/C/G)-3'.

FEATURES

source

Location/Qualifiers
 1. 661
 /organism="Glycine max"
 /db_xref="taxon:3847"
 /clone="Gm-r1070-5112"
 /clone_lib="Gm-r1070"
 /note="The library Gm-r1070 is a sequence-driven, rerecked set of 9,216 clones selected from cDNA libraries from various tissues and stages of development of soybean that represent 2,639 sequences from immature cotyledons, 1,770 from immature seed coats, 3,938 from flowers, and 869 from young pods. The 5' ESTs of the source clones from the different libraries were used to select singletons, or a representative of each contig, which were rerecked to form library Gm-r1070. The cDNA clones of the rerecked Gm-r1070 library were then sequenced at the 3' end. The contig analysis to select unique genes was performed by the laboratory of Ernest Retzel, Center for Computational Genomics and Bioinformatics, University of Minnesota, http://www.cbc.umn.edu/ResearchProjects/soybean/index.html. Rerecking was performed by Genome Systems, St. Louis, http://www.genomesystems.com, and 3' sequencing by the Keck Center for Comparative and Functional Genomics, University of Illinois, http://www.life.uiuc.edu/biotech/keck.html. Note: The corresponding 5' EST from each clone in the Gm-r1070 library is listed in the 'OTHER EST' field. The detailed information on the source library for each clone can also be obtained by referring to the Genome Systems clone ID of the original cDNA library that is also listed under 'OTHER EST'."

BASE COUNT 190 a 142 c 132 g 209 t 8 others
 ORIGIN

Query Match 60.0%; Score 15; DB 140; Length 681;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 gatctgctgtttgaa 16
 Db 490 GATCTGCTGTTTGA 476

RESULT 43
 AM128741/c

LOCUS AM128741 748 bp mRNA EST 25-OCT-1999
 DEFINITION fe36h03.x1 zebrafish Washu MPING EST Danio rerio cDNA 5' similar to
 TR:0920X1 Q920X1 APOPTOSIS-INDUCING FACTOR A1f.; mRNA sequence.
 AM128741
 AM128741.1 GI:6116645

ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

zebrafish.
 Danio rerio
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
 Cypriniformes; Cyprinidae; Rasbora; Danio.
 1 (bases 1 to 748)
 Clark, M., Johnson, S.L., Lehrach, H., Lee, R., Li, F., Marra, M., Eddy
 S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood
 K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Peterson, B.,
 Waller, T., Gibbons, M., Pape, D., Harvey, N., Schur, R., Ritter, E.,
 Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R.
 and Wilson, R.
 Washu Zebrafish EST Project 1998
 Unpublished (1998)
 Other ESTs: fe36h03.x1

TITLE

JOURNAL

COMMENT
 Contact: Stephen L. Johnson
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800
 Fax: 314 286 1810
 Email: zbrfish@watson.wustl.edu
 CDNA Library Preparation: Matthew Clark. CDNA Library Arrayed by:
 Matthew Clark. DNA Sequencing by: Washington University Genome
 Sequencing Center Clone Distribution: Genome Systems, St. Louis,
 Missouri (web address: www.genomesystems.com) (email contact:
info@genomesystems.com) and Research Genetics, Huntsville, Alabama
 (web address: www.resgen.com) (email contact: info@resgen.com) and
 ResourcenCentrumPrimaatenbank, Berlin, Germany (web address:
www.rzpd.de)

Seq primer: T3 ET from Amersham
 High quality sequence stop: 487.
 Location/Qualifiers

1.748

/organism="Danio rerio"
 /db_xref="taxon:7955"
 /clone_lib="Zebrafish Washu MPMG EST"
 /sex="mixed"
 /tissue="embryo"
 /lab_host="XLI-blue MRF"
 /note="Vector: pSPORT1. Site 1: NotI; Site 2: SalI; 1st
 strand cDNA was primed with a Not I oligo (drr15 primer
 15' pGACGATGCTGATGCGGCGCCCTTTTCTTTTCTTTT3 1);
 double-stranded cDNA was ligated into the Not I and Sal I
 digested with Not I and cloned into the Not I and Sal I
 sites of the pSPORT1 vector (BRL). Library was constructed
 by Matthew Clark (Lehrach lab; ICRF, London and Max Planck
 Institut fuer Molekulare Genetik, Berlin). cDNAs for EST
 analysis were selected following oligonucleotide
 hybridization fingerprinting of arrayed clones from
 zebrafish late somitogenesis (26 ss), adult liver or
 embryonic shield stage (5.6 h) libraries. Fingerprint
 data were used to computationally cluster cDNAs, and a
 single cDNA from each cluster was chosen for sequencing.
 In some cases multiple members of the same cluster were
 sequenced to assess clustering parameters or single clones
 were sequenced additional times to assess quality
 control."

BASE COUNT 183 a 186 c 197 g 181 t 1 others
 ORIGIN

Query Match 60.0%; Score 15; DB 112; Length 748;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 cgtttgaagcgag 22
 |||||||||||||
 Db 415 CTTGTAAGCGCAG 401

RESULT 44
 LOCUS 752 bp DNA GSS 14-MAY-2000
 DEFINITION Tetraodon nigroviridis genome survey sequence PUC-Or1 end of clone
 152M04 of library G from Tetraodon nigroviridis, genomic survey
 sequence.
 ACCESSION AL206315.1 GI:7865134
 VERSION AL206315
 KEYWORDS GSS; genome survey sequence.
 SOURCE Tetraodon nigroviridis.
 ORGANISM Tetraodon nigroviridis.
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
 Tetraodontidae; Tetraodon.
 1 (bases 1 to 752)
 Roest-Crollius, H., Jallion, O., Dasilva, C., Fitzames, C., Fisher, C.,
 Bouneau, L., Billault, A., Queller, F., Saurin, W., Bernot, A. and
 Weissenbach, J.
 Characterization and repeat analysis of the compact genome of the

JOURNAL
 REFERENCE
 AUTHORS
 TITLE

freshwater pufferfish Tetraodon nigroviridis
 Unpublished
 2 (bases 1 to 752)
 Roest-Crollius, H., Jallion, O., Dasilva, C., Bouneau, L., Fisher, C.,
 Bernot, A., Fitzames, C., Mincker, P., Brotlier, P., Queller, F.,
 Saurin, W. and Weissenbach, J.
 Human gene number estimate provided by genome wide analysis using
 Tetraodon nigroviridis DNA sequence
 Unpublished
 3 (bases 1 to 752)
 Genoscope.
 Direct Submission
 Submitted (12-APR-2000) to the EMBL/Genbank/DBJ databases
 This sequence is a single read and was generated as part of a large
 scale clone-end sequencing project of the Tetraodon nigroviridis
 genome. For more information, please take a look at
<http://www.genoscope.cns.fr/Tetraodon>.

FEATURES
 source
 1.752
 /organism="Tetraodon nigroviridis"
 /db_xref="taxon:99883"
 /clone_lib="G"
 /clone_lib="G"
 /note="Genoscope sequence ID : C0AG152DG02SP1-end :
 PUC-Or1"

BASE COUNT 138 a 199 c 238 g 163 t 14 others
 ORIGIN

Query Match 60.0%; Score 15; DB 220; Length 752;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 ttgaagcgagcgag 25
 |||||||||||||
 Db 698 TTTGTAAGCGCAG 684

RESULT 45
 LOCUS 846 bp mRNA EST 14-MAR-2001
 DEFINITION 602501093F1 NIH_MGC-75 Homo sapiens cDNA clone IMAGE:4614647 5',
 mRNA sequence.
 ACCESSION BC428569
 VERSION BC428569
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 846)
 NIH-MGC <http://imgc.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: CLONTECH Laboratories, Inc.
 CDNA Library Preparation: CLONTECH Laboratories, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
 Plate: LINC1366 row: j column: 24
 High quality sequence stop: 678.
 Location/Qualifiers

FEATURES
 source
 1.846
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4614647"
 /clone_lib="NIH_MGC_75"
 /lab_host="DH10B (TI phage-resistant)"
 /note="Organ: kidney; Vector: pDNR-LIB (Clontech); Site_1:"

Sfil (ggcgccctcgcc); Site 2: Sfil (ggccattatggc); 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGGCCGCGCGGACATG-dt(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.65 kb (range 0.5-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH-MGC Library."

BASE COUNT 235 a 214 g 187 t

ORIGIN

Query Match 60.0%; Score 15; DB 153; Length 846;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 gatctgctgtttgaa 16
 |||||
 Db 732 GATCTGCTGTTTGAA 718

Search completed: October 9, 2001, 15:15:28
 Job time: 13650 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 15:45:19 ; Search time 1670.83 Seconds
(without alignments)
231.438 Million cell updates/sec

Title: US-09-396-196f-5
Perfect score: 25
Sequence: 1 gcgcgcgcgcgcgcgcgcgcgcgcgcgcgcatt 25

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1344157 seqs, 7733874588 residues

Word size: 9

Total number of hits satisfying chosen parameters: 220030

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database:

GenBml: *
1: gb_ba1: *
2: gb_ba2: *
3: gb_ba3: *
4: gb_in1: *
5: gb_in2: *
6: gb_in3: *
7: gb_om: *
8: gb_ov: *
9: gb_pat1: *
10: gb_pat2: *
11: gb_ph: *
12: gb_pl1: *
13: gb_pl2: *
14: gb_pl3: *
15: gb_pl4: *
16: em_ba1: *
17: em_ba2: *
18: em_fun: *
19: em_htgo_hum: *
20: em_htgo_inv: *
21: em_htgo_rod: *
22: em_htg_hum1: *
23: em_htg_hum2: *
24: em_htg_hum3: *
25: em_htg_hum4: *
26: em_htg_hum5: *
27: em_htg_hum6: *
28: em_htg_hum7: *
29: em_htg_hum8: *
30: em_htg_inv1: *
31: em_htg_inv2: *
32: em_htg_other: *
33: em_htg_rod: *
34: em_hum1: *
35: em_hum2: *
36: em_hum3: *
37: em_hum4: *
38: em_hum5: *
39: em_hum6: *
40: em_hum7: *
41: em_in: *
42: em_om: *
43: em_or: *

44: em_ov: *
45: em_pat: *
46: em_ph: *
47: em_pl: *
48: em_ro: *
49: em_sts: *
50: em_sy: *
51: em_un: *
52: em_v1: *
53: gb_sts1: *
54: gb_sts2: *
55: gb_sts3: *
56: gb_sy: *
57: gb_un: *
58: gb_v11: *
59: gb_v12: *
60: gb_htg1: *
61: gb_htg2: *
62: gb_htg3: *
63: gb_htg4: *
64: gb_htg5: *
65: gb_htg6: *
66: gb_htg7: *
67: gb_htg8: *
68: gb_htg9: *
69: gb_htg10: *
70: gb_htg11: *
71: gb_htg12: *
72: gb_htg13: *
73: gb_htg14: *
74: gb_htg15: *
75: gb_htg16: *
76: gb_htg17: *
77: gb_htg18: *
78: gb_htg19: *
79: gb_htg20: *
80: gb_htg21: *
81: gb_htg22: *
82: gb_htg23: *
83: gb_htg24: *
84: gb_htg25: *
85: gb_pr1: *
86: gb_pr2: *
87: gb_pr3: *
88: gb_pr4: *
89: gb_pr5: *
90: gb_pr6: *
91: gb_pr7: *
92: gb_pr8: *
93: gb_pr9: *
94: gb_ro1: *
95: gb_ro2: *
96: gb_in4: *
97: gb_pr10: *
98: em_ba3: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	length	DB	ID	Description
1	25	100.0	1041	9	AR029499	AR029499 Sequence
2	25	100.0	1041	9	AR034916	AR034916 Sequence
3	25	100.0	1084	9	A11530	A11530 B10B gene o
4	25	100.0	5793	2	ECOB10	J04423 E.coli 7.8-
5	25	100.0	5872	9	A38246	A38246 Sequence 1
6	25	100.0	5872	9	A38251	A38251 Sequence 6
7	25	100.0	5872	9	A93674	A93674 Sequence 1
8	25	100.0	5872	9	A93679	A93679 Sequence 6

```

9      25 100.0 5872 9 AR101809 Sequence
10     25 100.0 5872 9 AR101810 Sequence
11     25 100.0 11022 1 AE000180
12     25 100.0 13501 1 AE005258
13     25 100.0 297816 2 AP002553
14     22 88.0 1121 10 E00893
15     20 80.0 2923 3 MAV250020
16     19 76.0 3603 3 PSEBHP1A
17     19 76.0 44201 3 MTVCY180
18     18 72.0 37500 73 AC068068
19     18 72.0 38000 61 AC010079
20     18 72.0 42000 77 AC087836
21     17 68.0 12891 1 AE004192
22     16 64.0 4674 89 AK024397
23     16 64.0 10290 1 AE005685
24     16 64.0 13350 3 MTVC031
25     16 64.0 23740 3 MTVC026
26     16 64.0 34902 3 POL245436
27     16 64.0 49587 6 DMBH48C10
28     16 64.0 60573 65 AC018267
29     16 64.0 82289 13 ATAC009895
30     16 64.0 115955 62 AC011732
31     16 64.0 177380 85 AC002457
32     16 64.0 180700 76 AC084136
33     16 64.0 289090 4 AE003424
34     15 60.0 893 53 CENS01JLR
35     15 60.0 978 53 CENS06ZHL
36     15 60.0 3069 15 AF219383
37     15 60.0 3369 13 AF274670
38     15 60.0 3369 13 AF274671
39     15 60.0 13916 3 MTCY9CA
40     15 60.0 35209 1 AF065159
41     15 60.0 35573 4 AC009603
42     15 60.0 41193 4 AC013257
43     15 60.0 44545 61 AC010753
44     15 60.0 55307 61 AC014684
45     15 60.0 73691 63 AC014903

```

ALIGNMENTS

```

RESULT 1
LOCUS AR029499 1041 bp DNA
DEFINITION Sequence 7 from patent US 5859335.
ACCESSION AR029499
VERSION AR029499.1 GI:5941472
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1041)
AUTHORS Patton,D.Andrew.
TITLE Enhanced biotin biosynthesis in plant tissue
JOURNAL Patent: US 5859335-A 7 12-JAN-1999;
FEATURES
source 1. 1041
Location/Qualifiers
BASE COUNT 262 a 273 c 305 g 201 t
ORIGIN

```

```

Query Match 100.0%; Score 25; DB 9; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.0026;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 ggcgcagcagtgatgcgcagcatt 25
|||||
Db 79 GCGCAGCAGGTGCATCCGACGATT 103

```

RESULT 2

```

AR034916
LOCUS AR034916 1041 bp DNA
DEFINITION Sequence 7 from patent US 5869719.
ACCESSION AR034916
VERSION AR034916.1 GI:5950521
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1041)
AUTHORS Patton,D.A.
TITLE Transgenic plants having increased biotin content
JOURNAL Patent: US 5869719-A 7 09-FEB-1999;
FEATURES
source 1. 1041
Location/Qualifiers
BASE COUNT 262 a 273 c 305 g 201 t
ORIGIN

```

```

Query Match 100.0%; Score 25; DB 9; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.0026;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 ggcgcagcagtgatgcgcagcatt 25
|||||
Db 79 GCGCAGCAGGTGCATCCGACGATT 103

```

```

RESULT 3
LOCUS A11530 1084 bp DNA
DEFINITION Biot gene of E.coli with primers.
ACCESSION A11530
VERSION A11530.1 GI:490218
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Escherichia coli.
REFERENCE 1 (bases 1 to 1084)
AUTHORS Patent: GB 2216530-A 16 11-OCT-1989;
JOURNAL Location/Qualifiers
FEATURES
source 1. 1084
gene /organism="Escherichia coli"
24. 1064 /db_xref="taxon:562"
24. 1064 /gene="BIOB"
24. 1064 /gene="BIOB"
24. 1064 /codon_start=1
24. 1064 /protein_id="CAA00965.1"
24. 1064 /db_xref="GI:490218"
24. 1064 /db_xref="SWISS-PROT:P12996"
24. 1064 /translation="MAHRRRMTLSVTELEFEKPLDILFEAQVHRHFFDPQVQVST
LSTKGCAPEDCKTCPOSSRYKTGLEARLMEVQVLESARAKAAGSTRGGAAM
KNHERDMPTLEDMVGVKAWGLEACMTGLTSEQAORLANAGLDYNNHLDTPSP
YGNITTYRTYQERLDTLEKVRDAGIKVCSGGIVGIGETVKDRAGLLDLANLPPEP
VPINMLVKVGTPLADNDVDVDAFDEIRTAIVARIAMPISYVLSAGREONBOQAAC
FMAGANSIFYGCKLITTPNREBDKDLQFRKRLGLNPQQTAVLAGNBQQQLBQAALMT
PVDDEVYNAAL"
CDS

```

```

BASE COUNT 271 a 286 c 318 g 209 t
ORIGIN

```

```

Query Match 100.0%; Score 25; DB 9; Length 1084;
Best Local Similarity 100.0%; Pred. No. 0.0026;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 ggcgcagcagtgatgcgcagcatt 25
|||||

```

Db 102 GCCGACGAGTGCATGCCAGCAT 126

RESULT 4

LOCUS EC0B10

DEFINITION E.coli 7, 8-diamino-pelargonic acid (bioa), biotin synthetase (biob), 7-keto-8-amino-pelargonic acid synthetase (biof), bioc protein, and dehydrobiotin synthetase (biocd), complete cds.

ACCESSION J04423

VERSION J04423.1

KEYWORDS GI:145422

SOURCE 7, 8-diamino-pelargonic acid aminotransferase; 7-keto-8-amino-pelargonic acid synthetase; bioc gene; biob gene; bioc gene; biob gene; bioc gene; biotin synthetase; dehydrobiotin synthetase.

ORGANISM Escherichia coli (strain K-12) DNA.

REFERENCE 1 (bases 1 to 5793)

AUTHORS Otsuka, A.J., Buoncrisiani, M.R., Howard, P.K., Flamm, J. and Johnson, O.

TITLE The Escherichia coli biotin biosynthetic enzyme sequences predicted

JOURNAL J. Biol. Chem. 263, 19577-19585 (1988)

COMMENT Draft entry and computer-readable sequence [1] kindly submitted by A.Otsuka, 09-NOV-1988

FEATURES

source

1. 5793

/organism="Escherichia coli"

/strain="K-12"

/db_xref="taxon:562"

complement(98..574)

/note="ORF 1"

/codon_start=1

/transl_table=1

/protein_id="AA023513.1"

/db_xref="GI:145168"

/translation="MKLISNDLRDGLKPHRNVGNGYGDGNTISPHLAWDDVPAGTKSPVVCYDPDAPGSGMMHVVVNI.PADTRVLPQGFSGSLVAMPDGLVLRTPDGKRGVDAAPKPKGERTHRYFTYVHALDIERIDVDEGASAAVGFNVHFSLSASITAMFS"

complement(633..1925)

/gene="bioA"

complement(633..1925)

/gene="bioa"

/codon_start=1

/transl_table=1

/product="7, 8-diamino-pelargonic acid aminotransferase"

/protein_id="AA023514.1"

/db_xref="GI:1457106"

/translation="MTTDLAFDQRIHWPYTSPLPYVVSABCECLISDGRKLVDSGSMWAAIHGYNHNPOLNAAKMSQIDAMSHVFGGITHAPALICRKLVAMSGRNALCEVFLADSGSVAVEVAMKALQYOKAKGEARQFLFERNGYHGDTPGASCDPDN SMHSLKQYLPENLAPAPASPDGEMDPMWGFARLMAHREIAVILTEPIVOGAGGRMTHPMLKRIKICDRGILLIADETIATGRTGKILACAEHAIAPILICIGKALTGCTMTLSATLTREVAFTISNGECMHPPTMGNPLCAANAASLALLESQDMOQQVADIEVQLRQDLAPARDAEMVADVRLGATGVETTHPVNMAALQKFPVEGQWIRPFGLIYMPPIYLLPQOLRLTAANRAVNODETFEQ"

2012..3052

/gene="biob"

2012..3052

/gene="biob"

/codon_start=1

/transl_table=1

/product="biotin synthetase"

/protein_id="AA023515.1"

/db_xref="GI:145425"

/translation="MAHRPMTLSOYTELFEKPLDLLFEAOVHRHDPPOVOYSTLSTKTCAPEDCKYCKQTSRYKTLFAPLMEVEVLSSARAKAGSTRCKMAAMKNPHERPYLEOVGVKAMGLACHTLCTLSESOAOLNAGLIDYVNHNDTPSEFYGNITRTYQERLDTLEKVDAGIKVCSGIVGLGELTYKDRAGILLIOLANPTPESVPIMLVKVGKTLADNDVDADFDFIKTIAVARIMPTSTVRLSAGREONEDTOAMC

gene

3049..4203

/gene="bioc"

3049..4203

/gene="bioc"

/codon_start=1

/transl_table=1

/product="7-keto-8-amino pelargonic acid synthetase"

/protein_id="AA023516.1"

/db_xref="GI:145426"

/translation="MSWDERINAAALDARRADALRRRYPAVAGGRMLVADROYLNFSSNDYGLSHHPOLIRAMOGAEOFGISGSGHVSYSVHQALEELALMLTYSRALFTISGFANNOAVIAMAAMAKERRIADRLSHSLASLSPSOULRRFANDVTHLARLHAPROGOMVTEGVFMSDDSPALETIOQVTOOHNGMLMVDPDAGTGVIGQSGSCWLOKYPRELLVYTRKGFVSGAAGVCSVAVYLLQFARHLIYSTMPDPAQALRASLAVIRSDGARRREKIALITTRFRAGVODLPFLADSCAIOPLIVGNSRALOLAEKLRQGCWVTARIPPTVAGTARLRLTLTAHHEMODIRLEVLHONG"

4190..4945

/gene="bioc"

4190..4945

/gene="bioc"

/codon_start=1

/transl_table=1

/protein_id="AA023517.1"

/db_xref="GI:145427"

/translation="MATVKNQALIAAFGRRAAHYEQHADLRQSDADALLAMPORRYTHVLADGCGPCGMSHMRERHAOYVATLDSPLVYVQAROKDADHYLAGDIESLPLATAFTDILAMSLAVQMGKNSLTALRELYRVRRGVAVFTLVQGLSPRRHQAQVADBERP HANRFLPPEDEISSLNGVYHQHHPOTITMPEDALSMKSLKIGATFHLHEGRDPRILTRSQLRLQIALAMPQOGGRPLTYHLFLGVARE"

4938..5597

/gene="biob"

4938..5597

/gene="biob"

/codon_start=1

/transl_table=1

/product="dehydrobiotin synthetase"

/protein_id="AA023518.1"

/db_xref="GI:145428"

/translation="MSKRYFVYGTDFEYKTVASCALLQAKRAGRYTGYKRVASGSETPRGLNSQALMORNSLQDLQYATVNPYTFEAPSPHIISSQDEKRPESLYMSAGLHALDQADWVLYVEGAGCGMPLSDTFEPADWVVOEDLPVLLVGVYGLGGINHAMTLA OVIOHAGLTLGAVANDVTPPGKKAHEKMTLTLMIRRCJERSPLQKIQKKROES"

BASE COUNT 1363 a 1631 g 1245 t

ORIGIN 4626 bp upstream of HpaI site; 18 min on K-12 map.

Query Match 100.0%; Score 25; Length 5793;

Best Local Similarity 100.0%; Pred. No. 0.0019;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcgacgagtgatcgccagcatt 25

Db 2090 gccgacgagtgatcgccagcatt 2114

RESULT 5

LOCUS A38246

DEFINITION Sequence 1 from Patent WO9408023.

ACCESSION A38246

VERSION A38246.1

KEYWORDS GI:2294844

SOURCE Escherichia coli.

ORGANISM Escherichia coli.

REFERENCE 1 (bases 1 to 5872)

AUTHORS Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.

TITLE BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN

JOURNAL Patent: WO 9408023-A 14-Apr-1994;

PAT 05-MAR-1997

```

COMMENT
LONZA AG (CH)
Other publication PL 308301 950724
Other publication CA 2145400 940414
Other publication AU 4820293 940426
Other publication HU 71781 960228
Other publication SK 42095 951108
Other publication CZ 9500809 950913
Other publication FI 951547 950331
Other publication JP 8501694T 960227.
location/Qualifiers
1. 5872
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/clone="PBO30A-15/9"
1. .96
/function="PROMOTER PTAC"
/evidence=experimental
23. .28
/standard_name="PROMOTER PTAC"
/evidence=experimental
105. .119
/standard_name="BIOB RBS NO.9"
/evidence=experimental
117. .1157
/standard_name="BIOB"
/clone="BIOB"
/codon_start=1
/transl_table=11
/number=1
/evidence=experimental
/product="BIOTIN SYNTHASE"
/protein_id="CAA02325.1"
/db_xref="GI:2294845"
/translation="MAHPRKWTLSQVTELEFKPLLDLLEAQQVHROHPDROVQVST
LISITGACPECKTCPOSSRYKGTGLEAEERLVEQVLESARAKAGSTRPGMAAM
KNHPRDPELEQWQVGVKAMGLACMTLGLTSEQARLANGLDYNINLTPSEF
KNIITTRPYRELDLEKVDAGIKVCSGIVGLGETVDRAGLILLOLNLPTPES
VPIMLVVKGTPLADNDVDAPFRTIIVARIIMPTSVYRISAGREONEDPTQMC
FMAGANSIFYGCKLITTPNPEDKDIDLFRLGLNPQATVLAGDNBQDRLOALMT
PDDEYVNAAL"
2284. .2297
/standard_name="BIOC RBS"
2295. .3050
/standard_name="BIOC"
2295. .3050
/standard_name="BIOC"
/function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
/codon_start=1
/transl_table=11
/number=3
/product="PROTEIN"
/protein_id="CAA02326.1"
/db_xref="GI:2294846"
/translation="MATVNRQAIAAAGRAAHYEQHADLQROSADALIAMLPORKYT
HYLDAGCGPEWMSRHRERAAQVATLALDSPMLVQARQDADHLAGDTESTPLATA
TFDLASNLAVQCNLSTALRELYVVRPKGVATTTVOGSLPEIIOARQAVDERP
HANRFLPDEIEQLNGVYOHIOITLIMPDAISAMSLKGLGATHLHREGDRILL
TRSOLORLQAWVQGGRYPLTHFLGVIANE"
3742. .3752
/standard_name="BIOA RBS"
3750. .5039
/standard_name="BIOA"
3750. .5039
/standard_name="BIOA"
/standard_name="S-ADENOSYL-L-METHIONINE:
8-AMINO-7-OKONONANATE AMINOTRANSF."
/RC_number="2.6.1.62"
/codon_start=1
/transl_table=11
/number=5
/organism="Escherichia coli"

FEATURES
source
/evidence=experimental
/product="DAPA SYNTHASE"
/protein_id="CAA02327.1"
/db_xref="GI:2294847"
/translation="MTTDLAFDORHIMPTYSMTSPLPVYPVSAEGCELLISNGR
LVQMSWMAAIIHGNHNPOLNAKMSQIDAMSHVWFGGITTAPELCEKRLVAMTPOR
LECVFLADSGSVAVEVAMKALQVQAKGAEARQRLTRNGHGTGFCMSVCDPNS
MHLMLGYPENLFAPQSRNDGENDENDVGRRLMAAHHEIAVITIEPIVQAG
GMRVPEWMLKRIKICDREGILLIADETATGFRGKLFACERHETIAPDILGKAL
TGSTMTLSATLTITREVAETISNGEAGCEFHGPTFMGNPLACAAANASIALIESGDMO
QVADIEVLREQAPARDAEMVADVRLGALIGVETTPVNMALOKFVQGWITP
EKLIIYIMPEYIILPQOLRLTAANRAVDETFCQ"
5088. .5100
/standard_name="ORF1 RBS"
5098. .5574
/standard_name="ORF1"
5098. .5574
/standard_name="ORF1"
5098. .5574
/standard_name="ORF1"
/function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
/codon_start=1
/transl_table=11
/number=6
/evidence=experimental
/product="PROTEIN"
/protein_id="CAA02328.1"
/db_xref="GI:2294848"
/translation="MKLISNDLRDGDKLPHRYVNGNGYGDGINSPLHLDVPACTK
SEVVCYDPAAPGTGSGMHWYVYVLPADTRVLPDGRSGGLVAMPDGLQTRTPGRTG
YDGAAPKRGTHRYITFVHALDTRIDVDSAGAMGVFVHFHSLASISITAMFS"
5583. .5605
5583. .5644
/standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
TERMINATOR"
BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN

Query Match
Best Local Similarity 100.0%; Score 25; DB 9; Length 5872;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ggcgcagcagtcctcgcgcagcatt 25
|||||
Db 195 gccacagcagtcctcgcgcagcatt 219

RESULT 6
A38251 5872 bp DNA PAT 05-MAR-1997
LOCUS
DEFINITION Sequence 6 from Patent WO9408023.
ACCESSION A38251
VERSION A38251.1 GI:2294849
KEYWORDS
SOURCE
ORGANISM
Escherichia coli.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE
1 (bases 1 to 5872)
AUTHORS Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
TITLE BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
JOURNAL Patent: WO 9408023-A 6 14-APR-1994;
LONZA AG (CH)
COMMENT
Other publication PL 308301 950724
Other publication CA 2145400 940414
Other publication AU 4820293 940426
Other publication HU 71781 960228
Other publication SK 42095 951108
Other publication CZ 9500809 950913
Other publication FI 951547 950331
Other publication JP 8501694T 960227.
location/Qualifiers
1. 5872
/organism="Escherichia coli"

```



```

RBS      5088..5100      /standard_name="ORF1 RBS"
gene      5098..5574      /gene="ORF1"
CDS       5098..5574      /function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
                        /codon_start=1
                        /transl_table=11
                        /evidence=experimental
                        /product="PROTEIN"
                        /protein_id="CAB69593.1"
                        /db_xref="GI:6741866"
stem_loop /translation="MKLISNDLRDCKLPHRIVFNGMGYDGNISPHLAMPDVPAGTK
terminator YDCAAPKGGTTHRYIFTVALDIERIDVDEGASGAMVGFNFHSLASASITAMS"
                        5583..5605
                        5583..5644
                        /standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN
Query Match      100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.0019;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ggcgcagcagtgcatgccagcatt 25
Db 195 GCCGACGAGTGTCATGCCAGCATT 219

RESULT 8
LOCUS      A93679      5872 bp      DNA
DEFINITION A93679      Sequence 6 from Patent EP0798384.
ACCESSION  A93679
VERSION    A93679.1 GI:6741867
KEYWORDS
SOURCE     Escherichia coli.
ORGANISM   Escherichia coli.
REFERENCE  1 (bases 1 to 5872)
AUTHORS   Birch,O. and Brass,J.
TITLE     Biotechnological method of producing biotin
JOURNAL   Patent: EP 0798384-A 6 01-OCT-1997;
          LONZA AG (CH)
FEATURES   location/Qualifiers
Source     1..5872
           /organism="Escherichia coli"
           /strain="DSM498"
           /db_xref="taxon:562"
           /clone="PBO30A15-9"
           1141..1156
           /standard_name="BIOP RBS"
           1154..2308
           /gene="BIOP"
           1154..2308
           /gene="BIOP"
           /standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
           /EC_number="2.3.1.47"
           /codon_start=1
           /transl_table=11
           /evidence=experimental
           /product="KAPA SYNTHASE"
           /protein_id="CAB69594.1"
           /db_xref="GI:6741868"
           /translation="MSMCKEINAAIDARRADALRRRYPVAGAGRWLVADROYLNF
SSNDYIGLSHHFPIIRAWOOGAEFGISGSGHYGVVYHQAIEELALMLGYSLRA
LFTISFPAANOAVIAAMAKEDRIADRLSHSLLEASTSPSLRRRAHNDVYHLAR

```

```

RBS      3043..3753      /standard_name="BIOD RBS"
gene      3043..3753      /gene="BIOD"
CDS       3043..3753      /standard_name="BIOD RBS"
                        /codon_start=1
                        /transl_table=11
                        /evidence=experimental
                        /product="DTB SYNTHASE"
                        /protein_id="CAB69595.1"
                        /db_xref="GI:6741869"
                        /translation="MSKRYFVGTDEYGTIVASCALLOAKAGYRFAGYRPAVS
EKTEPEGLRNSDALALORNSLQIDVATNPYTPAEPSPHISIQEGRPTESLVSAG
LRALBOQADWVIVGAGGMPETLSPTTFPADWYTOBOLPYLVGVGLGINHAMLTA
OYIOHAGLITLACWANDVTPPEKRAEYMTITLTMIPAPLGLIPLAENPENAAATGK
YINLAPVDASTIGFTSRL"
BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN
Query Match      100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.0019;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ggcgcagcagtgcatgccagcatt 25
Db 195 GCCGACGAGTGTCATGCCAGCATT 219

RESULT 9
LOCUS      AR101809      5872 bp      DNA
DEFINITION AR101809      Sequence 1 from patent US 6083712.
ACCESSION  AR101809
VERSION    AR101809.1 GI:12812607
KEYWORDS
SOURCE     Unknown.
ORGANISM   Unknown.
REFERENCE  1 (bases 1 to 5872)
AUTHORS   Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
TITLE     Biotechnological method of producing biotin
JOURNAL   Patent: US 6083712-A 1 04-JUL-2000;
          Location/Qualifiers
Source     1..5872
           /organism="unknown"
BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN
Query Match      100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.0019;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ggcgcagcagtgcatgccagcatt 25
Db 195 GCCGACGAGTGTCATGCCAGCATT 219

RESULT 10
LOCUS      AR101810      5872 bp      DNA
DEFINITION AR101810      Sequence 6 from patent US 6083712.
ACCESSION  AR101810
VERSION    AR101810.1 GI:12812608
KEYWORDS

```

SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 5872)
AUTHORS Birch, O., Brass, J., Fuhrmann, M., and Shaw, N.
TITLE Biotechnological method of producing biotin
JOURNAL Patent: US 6083712-A 6 04-Jul-2000;
FEATURES
Location/Qualifiers
source
1. 5872
/organism="unknown"
BASE COUNT 1318 a 1552 c 1695 g 1307 t
ORIGIN
Query Match 100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.0019;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gcgcagcagtgatgcgcagcatt 25
Db 195 GCGCAGCAGTGATGCCAGCAT 219
RESULT 11
AE000180 11022 bp DNA BCT 01-DEC-2000
LOCUS Escherichia coli K12 MG1655 section 70 of 400 of the complete
DEFINITION genome.
ACCESSION AE000180 U00096
VERSION AE000180.1 GI:1786988
KEYWORDS
SOURCE ORGANISM
Escherichia coli K12.
Escherichia coli K12.
Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
1 (bases 1 to 11022)
Blattner, F.R., Plunkett, G. III, Bloch, C.A., Perna, N.T., Burland, V.,
Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F.,
Gregor, J.J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J.,
Mau, B. and Shao, Y.
The complete genome sequence of Escherichia coli K-12
Science 277 (5331), 1453-1474 (1997)
TITLE
JOURNAL
PUBMED 97426617
9278503
2 (bases 1 to 11022)
Blattner, F.R.
Direct Submission
Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecolig@genetics.wisc.edu Phone: 608-262-2534 Fax:
608-263-7459
3 (bases 1 to 11022)
Blattner, F.R.
Direct Submission
Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecolig@genetics.wisc.edu Phone: 608-262-2534 Fax:
608-263-7459
4 (bases 1 to 11022)
Plunkett, G. III.
Direct Submission
Submitted (13-OCT-1998) Laboratory of Genetics, University of
Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
This sequence was determined by the E. coli Genome Project at the
University of Wisconsin-Madison (Frederick R. Blattner, director).
Supported by NIH grants HG00301 and HG01428 (from the Human Genome
Project and NCHGR). The entire sequence was independently
determined from E. coli K12 strain MG1655. Predicted open reading
frames were determined using Genemark software, kindly supplied by
Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
30332 (e-mail: mark@ember.gatech.edu). Open reading frames that
have been correlated with genetic loci are being annotated with CG
Site Nos., unique ID nos. for the genes in the E. coli genetic

Stock Center (CGSC) database at Yale University, kindly supplied by
Mary Berlyn. A public version of the database is accessible at
(http://cgsc.biology.yale.edu). Annotation of the genome is an
ongoing task whose goal is to make the genome sequence more useful
by correlating it with other data. Comments to the authors are
appreciated. Updated information will be available at the E. coli
Genome Project's World Wide Web site
(http://www.genetics.wisc.edu). *** The E. coli K12 sequence and
its annotations are periodically updated; this is version M54. No
sequence changes. Annotation updates: updated gene identifications
and products: all new functional assignments courtesy of Monica
Riley; added promoters, protein binding sites, and repeated
sequences described in reference 1. The unique numeric identifiers
beginning with a lowercase 'b' assigned to each gene (protein- or
RNA-encoding) are now designated as gene synonyms instead of
labels. This should allow them to be searched for in Entrez as gene
names.
FEATURES
source
Location/Qualifiers
1. 11022
/organism="Escherichia coli K12"
/strain="K12"
/sub_strain="MG1655"
/db_xref="taxon:83333"
<1. 130
/note="REP (repetitive extragenic palindromic) element;
contains 4 REP sequences"
complement(147..1430)
/gene="ybhC"
/note="b0772"
complement(147..1430)
/gene="ybhC"
/function="orf; Not classified"
/note="f427: 98 pct identical to fragment YBHC_ECOLI
SM:P46130 (300 aa) but contains 127 additional C-terminal
residues"
/codon_start=1
/transl_table=1
/product="putative peptidase"
/protein_id="AAC73859.1"
/db_xref="GI:1786988"
/translation="MNFESVSRLLALAFGVTLTACSTPPDQRPDSOTAPETSSRPI
LSAEADNPDAHYEFLTPGAAMNPSPTLTPAODFVVGPAAGTGVTHTTIGQAVD
AAITKRTKROYIAVMPGEGYOGGVYPAAGGTLTGEGKPIDVKIGSLDGSAPA
DMRHVDVREGKTPMGKPAWYATYSCSKRSQSDSGVCSAVFSONNGLOLNTLENT
LGSDVDGNHPAAVALRTDGDVOVINNVNIIIGRQNTFPYVNSCYONLEFNRPRITLT
NSYLEGDVIVSGCAVFPDTEFRVNVNSTQDAIVFAFATISNITTYGFLVANSFEN
AFGDGVOLGRSLDVAANTNGOVYIRDSAINEGFNAPKPAWADAVISNRPAGNTGSDV
DNDEIQRNLNTFNRYRMEXYNNRGVCSKVVAAEAKK"
complement(1507..1535)
/note="Factor Sigma70; predicted +1 start at 806574"
complement(1582..2058)
/gene="ybhB"
/note="b0773"
complement(1582..2058)
/gene="ybhB"
/function="orf; Unknown"
/note="f158: 99 pct identical to YBHB_ECOLI SM: P12994"
/codon_start=1
/transl_table=1
/product="orf, hypothetical protein"
/protein_id="AAC73860.1"
/db_xref="GI:1786990"
/translation="MKTISNDLRGDKLPBRHVENFGMGYDGNISPHLADNDVPAGTK
/STVAAPKGGTETRYIFVYHADDIERIDVDEASGAMCFNHFHSLASASITAFS"
YDCAAPKGGTETRYIFVYHADDIERIDVDEASGAMCFNHFHSLASASITAFS"
complement(2108..2124)
/note="central position to predicted promoter:85.5"
/bound_molety="Rhas predicted site"
complement(2117..3406)
/gene="b10A"
/note="b0774"
complement(2117..3406)
/gene="b10A"

CDS

protein_bind

gene

/EC_number="2.6.1.62"
/function="enzyme; Biosynthesis of cofactors, carriers:
Biotin"
/note="f429; 100 pct identical to B10A_ECOLI SW: P12995"
/codon_start=1
/transl_table=11
/product="7,8-diaminopelargonic acid synthetase"
/protein_id="AAC73861.1"
/db_xref="GI:1786991"
/translation="MTTDLAFDPQRIHPIYMSPLPVPVSAEGCELLISDGR
LVGMSWMAAIGHYGNHPOLNAAKSDIDAMSHVMGGITHAPELCKIYAMTPOP
LECPFLADSGSVAEVAAMKALQYWAOKGEARQPLTFENGHCDPFGAMSCDDNS
MSLMKGYLPENLEFAPAPQSRMDGEWDERDMVGFALMAAHRHETAAVITYEYOGAC
GMRHYHEMLKRIKICDREGILLADIELATGEGRGKLPACENAFIAPDILCKAL
TGGMTLSATLTREVAFTISNGEACFMHGPFGNPLACAAASALIESGDMOO
QVADIEVQLEOLAPADAMVADVPVCAIVETTHPVMALAKFFVBOGWIRP
FGKILMPYIILPQOILRTAAVRAVQDETFPCQ"
complement(2193..2221)
/gene="bioA"
/note="factor Sigma70; predicted +1 start at 807260"
3411..3450
/note="central position to bioB promoter: -20"
/bound_moiety="BioB documented site"
3411..3450
/note="central position to predicted promoter: -20; genetic
evidence for the site"
/bound_moiety="BioB predicted site"
complement(3411..3450)
/note="central position to predicted promoter:50"
/bound_moiety="BioB predicted site"
complement(3411..3450)
/note="central position to bioA promoter:50"
3413..3441
/bound_moiety="BioB documented site"
808525
/note="factor Sigma70; promoter bioB; documented +1 at
complement(3447..3473)
/note="factor Sigma70; promoter bioA; documented +1 at
808515"
4533
/gene="bioB"
/note="b0775"
3493..4533
/gene="bioB"
/EC_number="2.8.1.-"
/function="enzyme; Biosynthesis of cofactors, carriers:
Biotin"
/note="o346; 99 pct identical to B10B_ECOLI SW: P12996"
/codon_start=1
/transl_table=11
/product="Biotin synthesis; sulfur insertion?"
/protein_id="AAC73862.1"
/db_xref="GI:1786992"
/translation="MAHRRWTLSQVTELEFEKPLDLEFAQOVHROHFPPOVST
LLSTKGACPEDCKYCPQSSRYTGLEARLMEVEQVLESARAKAAGSTRMGAAW
KNHERMPYLEQWQVQKAMGLEACMTGLTSESQALANAGLDYNNHNDPSPEP
YGNITTRTYQERLDTLEKVRDAGIKVGGSGIVGGEYKAGLILLOUANIPNPES
VPINLVKVGKPTLADNDVDADDFIRTIATVAVRIMPPSYVLSGRGEMQOTAMC
FMAGANSIFGCKPLFTTPNEEDKDLQFLRKGLNPOQTAVIAGNEQOORLEQALMT
PTDEYVNAALM"
4530..5684
/gene="bioF"
/note="b0776"
4530..5684
/gene="bioF"
/EC_number="2.3.1.47"
/function="enzyme; Biosynthesis of cofactors, carriers:
Biotin"
/note="o384; 100 pct identical to B10F_ECOLI SW: P12998"
/codon_start=1
/transl_table=11
/product="8-amino-7-oxononanoate synthase"
/protein_id="AAC73863.1"

Query Match 100.0%; Score 25; DB 1; Length 11022;
Best Local Similarity 100.0%; Pred. No. 0.0017;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 ggcgcagcaggtgcacgcgcagcatt 25
Db 3571 GCGCAGCAGCGATCCCGCAGCATT 3595
/db_xref="GI:1786993"
/translation="MSMOEITNAALDARRAADALRRRYPVAGAGRWLVADROYLNF
SSNDYIGLSIHHPQITIRAWQGAQGFIGSGGSHSVSVVQALAEELAEALGYSSRA
LFTISGFANQAVVIEAAMAKERIALDRISHALISLASISLSPQALRRRAHDVHTLAR
LLASPCGQOMVTEGVFSMDGSDAPLARIQOVTOOHNGMLMDVGDHGTGYTGGRG
SCWLQKVKPELAVTEFGKGFVSGAAVILCSSIVADYILQFARHLITYSTMPAPQAL
RESULT 12
AE005258
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Escherichia coli O157:H7 EDL933.
Escherichia coli O157:H7 EDL933.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
1 (bases 1 to 13501)
Perna,N.T., Plunkett,G., III, Burland,V., Mau,B., Glasner,J.D.,
Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
Postal,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
Grobeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamoudis,K.,
Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
Welch,R.A., and Blattner,F.R.
Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
Nature 409 (6819), 529-533 (2001)
21074935
PUBMED
TITLE
JOURNAL
MEDLINE
REFERENCE
Perna,N.T., Plunkett,G., III, Burland,V., Mau,B., Glasner,J.D.,
Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
Postal,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
Grobeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamoudis,K.,
Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
Welch,R.A., and Blattner,F.R.
Direct Submission
Submitted (22-OCT-2000) Laboratory of Genetics, University of
Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
1. 13501
Location/Qualifiers
/organism="Escherichia coli O157:H7 EDL933"
/strain="EDL933"
/serotype="O157:H7"
/db_xref="taxon:155864"
/note="enterohaemorrhagic"
<1..7576
/note="O-island #36: Region of the EDL933 chromosome not
homologous to E. coli K-12 MG1655; Cryptic prophage
CP-933K; Includes one copy of the 13 bp direct repeat that
flanks the prophage"
66..665
/gene="lomK"
/note="Z0981"
66..665
/gene="lomK"
/function="putative membrane; Other or unknown (Phage or
Prophage Related)"
/note="Residues 1 to 199 of 199 are 72.86 pct identical to
residues 1 to 199 of 199 from Genpept 118 :
g417532789[gb|AA063231.1|AF151091_2 (AF151091) lom
[prophage P-Eiba]"
/codon_start=1


```

/transl_table=11
/product="putative outer membrane protein of prophage
CP-933K"
/protein_id="AAG55137.1"
/db_xref="GI:12513752"
/translation="MKVCAILSLATLAVGVPANASEHOSTLSAGYLHASTDAPC
SDLDINGIKRYEFTDPLGLTFSYNADEQKTHSDTRHEDYVRNRSVYAG
PSVRVEMESAAMAGAVASRYSTFSGDYFRVTDNKRTHDVLGSDARYSNTSLAW
GAGVQNPTESAVDAYEXSSGSDMRTDGFIVGYKF"
/gene="20982"
/size="2041"
/function="putative structure; Structural component (Phage
or Prophage Related)"
/notes="Residues 164 to 440 of 440 are 68.79 pct identical
to residues 381 to 645 of 645 from GenPept 118 :
g1145854361gb|AAD25464.1|AF125520.59 (AF125520) putative
tail fiber protein [Bacteriophage 933W]"
/codon_start=1
/transl_table=11
/product="putative tail component of prophage CP-933K"
/protein_id="AAG55138.1"
/db_xref="GI:12513753"
/translation="MGMAVQISGVKDCAGKPIQNCITOLKARRNSTXVVNTVASE
NPDEAGRYSDVEYGYVTLVEGPPSHAGTISVEDSQGTINDFLGAMTEDDAR
PEALRRFEQVEEAARHAEARKNAGEACTSARNAGISAKASANAADVSAEDAR
SARQAESASAKSEASASSASASASASASASASASASASASASASASASASAS
KSTFKAREASASASASASASASASASASASASASASASASASASASASASAS
RQTPGKAGTASASASASASASASASASASASASASASASASASASASASASAS
GPOGPGKGTGAAPVATGPGKPGKPGKPGKPGKPGKPGKPGKPGKPGKPGKPGK
TGLTFDPRKATOVGMFQHLQVRFEDGFQWQDKGLDEVSGTGRTGE"
/size="2312"
/gene="20984"
/size="2004"
/function="orf; Other or unknown (Phage or Prophage
Related)"
/notes="Residues 1 to 102 of 102 are 98.03 pct identical to
residues 28 to 129 of 129 from GenPept 118 :
g1145854371gb|AAD25465.1|AF125520.60 (AF125520)
hypothetical protein [Bacteriophage 933W]"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55139.1"
/db_xref="GI:12513754"
/translation="MKWAVIOAELENDMNLIRKLMOISLGCCKHDDCENGSRSLTAOLR
LGPADLLESDENGILPEQDRVITQVVLADAKKQICQVVPQLILRADGTWIGMK"
/size="2489"
/gene="20985"
/size="2489"
/gene="20985"
/function="orf; Other or unknown (Phage or Prophage
Related)"
/notes="Residues 13 to 326 of 326 are 55.73 pct identical
to residues 18 to 331 of 336 from GenPept 118 :
g1169603671gb|AF33527.11 (AF170176) hypothetical protein
predicted by Glimmer [Salmonella typhimurium LT2]"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55140.1"
/db_xref="GI:12513755"
/translation="MLSPIRTTFHNSVNIYVSPQVTSFAGKEYELKVIDEPTPLF
LQEFENPERYKDEPVIYNTKQHYLDNVTNAARIESDRMIGFVVDGDFSVNOKTAFS
KLREDFEWMVITVEDVDSWDRKLSIYDIICPQRLRTEDKRDYELINLLEKEIR
ETSKRQDSLISYAKKRHHAAWDFPFRNALILKAGETRCGYNTRKHGISEGCEYLD
MDMILTKGLGITAPDGTSMHVDNRNVSNTENALITVNSNHPLLEGLSFMHSKVD
AHPIYDGLGKGVKKYFNPPLAHNYHFCDFLEFHNPIIMNTSOTYSSW"
/size="3530"
/gene="20986"
/size="3530"

```

```

CDS
3530..4522
/gene="20986"
/function="orf; Other or unknown (Phage or Prophage
Related)"
/notes="No significant matches"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55141.1"
/db_xref="GI:12513756"
/translation="MKIPSLQSNFENESAPAGYSAPAPNRAENAYADYVIDICKRIPL
SAADLSNVEYSVIRAVHDSRSLIDQHTVDNIGVTDALSRSOTFQDAVSYGHNEX
VHIGCIKYNEYELNEESSVKIDDIQSLTCLNELYDVGQEPFPEICEADENDEEY
VSPSYAPDQDSYEMPMOEGILHEIHHVTSSDPSGDSNTEIGPTIELARKVAQELG
MSVPQKGYAEPFERAHLRLRLNLRQAAHREHNERAEPERLGTISDRYASPDFT
ETSAVSNIGTGTQOHDPPGLAINQNDQANDQIDLYHAPYIFPFQGVKHNQ"
/size="5350"
/gene="20989"
/size="5350"
/function="orf; Other or unknown (Phage or Prophage
Related)"
/notes="No significant matches"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55142.1"
/db_xref="GI:12513757"
/translation="MLSPIVNIQCSMNSTFRNLTPDNRLSVYRDAVHSDNGAAY
KVANKRYAVATDNKFCVTRRESHGCTFLRLKLPKGEISRKLEWLNAPVSAAM
ERGIIVHSNRPDLPPVDYAPPELPSVYDNLSPGVNTIGGAAYVEDAEADATYAKM
FTTSQSNSEEVTSSEVRCENQYGGASAEKLYGNNGDILGRMKINGESLNIISLPAP
AEHRIYDMFRLKQKGLFVDTETVLYDRAKNFNPIDISSYVNSDRSSEQIMQ
SYHGKODLISVLSKI"
/size="6462"
/gene="20990"
/size="6462"
/gene="20990"
/function="orf; Other or unknown (Phage or Prophage
Related)"
/notes="No significant matches"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55143.1"
/db_xref="GI:12513758"
/translation="MRPTSLNVLVHOSRSSSSMDIDIESLVKASSQWIKNNPOLRF
OCTHNTYQOIEAALDKIGSTETGRVILNATESISRLKSEYVYIHLNSSLGVMARH
IDAENHNGTSDPFCNLAAYEPCGBCISYVDHATIVHEILHAFHNHNGRLAYES
SRAESQKYSPLLEEARVGLGAFSEVSENKFHEIGMPTRTSTPXSLALIHDDNT
VSLGQOVRHLHPL"
/size="6462"
/gene="20992"
/size="6462"
/gene="20992"
/complement(7668..8144)
/gene="yhbH"
/notes="yhbH"
/function="orf; Unknown function"
/notes="Residues 1 to 158 of 158 are 99.36 pct identical to
residues 1 to 158 of 158 from Escherichia coli K-12 Strain
MG1655: B0773"
/codon_start=1
/transl_table=11
/product="orf; hypothetical protein"
/protein_id="AAG55144.1"
/db_xref="GI:12513759"
/translation="MKLISNDLRDGLPKRHHVNGMGYDGNISPHLAWDDVPAGTK
SFVYCYDPPDAPTGSGMMHWYVVLPAIDTVLPQFGSGGLVAPDDGLQTRDFGKTG
YDGAAPVGEHRTITFYHALDVRIDVDEGASAMGFVNHSHSLASASITAMFS"
/complement(8203..9492)
/gene="20993"
/size="3530"

```

gene

gene

CDS

gene

CDS

gene

CDS

gene

CDS

CDS

gene

CDS

CDS


```

/product="putative cob(I)alamin adenosyltransferase"
/protein_id="BAB34179.1"
/db_xref="gi:13360215"
/translacion="MEARISTERNVORROOKLEQVDPYVAALIEKKGILLVFTGNGK
KSTAFGCVTRAVGHGCTVGAQYIKGQMDNCEYMLDPLGVFPIWGTGTWETQNK
QADIDAKKEWMSKEDRLADKRYDLVYDELTYMLAHYLDTEEVYIASLONRPAQOSV
IVTRGCHSOLQIKMADTVSEIRPVKHAFPNGIQAQGLDW"
complement(4332. .5984)
/gene="Ecs0757"
complement(4332. .5984)
/gene="Ecs0757"
/note="probable fumarate hydratase, similar to fumarate
hydratases e.g. fumarate hydratase class I, aerobic
(fumarase) - Escherichia coli
g1120598|sp|P00923|FUMA_ECOLI percent identity 68 in 545
aa"
/codon_start=1
/transl_table=1
/evidence=not_experimental
/product="putative fumarate hydratase"
/protein_id="BAB34180.1"
/db_xref="gi:13360216"
/translacion="MSKRFITWQEPFLONKGTETYLISDQHTVTELEDEEYIKAPI
A/LTLARAFYEASFPELSAHLDQVAVSLNDQOASNDKYVALQILRNMEVSKGVLP
NODQPTGATIVASKGOQIWTGNDGAEALSAGITISFOENLNR SQAPALDWTETKQV
TMLPAQIDISAVAGDEYHFLCVNKGGSANKMALYQPTKSLQDPEKTLFLBKMSL
GPAADPPHYAFVAGVGLSADQTLRLAKLTKKYDNLPTGSDNGOQAFNDIEKVL
EASQOFGAOGKGVGFYFANDIRVLRPRHGSGCPIAMALSCSDRNRIKAKINHGIML
EKLEHNPQYIPASIREENHOAHOLDLNPRLRVMODLARLPVTRVSLGCIYVAR
DIAHAKIARLDSKSGRSPREYIKKHITVYAGAKPRMAGCSGPTTGGMDQYIDTF
QAGASVLWLSKSGNSQOYDPACHKHGGFNIGSIGMAALIAOEYKSLRCLEYPELG
MEAVMMVEVENLPAILVDKGNFESFQDQHCASCAPAGH"
complement(6092. .7372)
/gene="Ecs0758"
complement(6092. .7372)
/gene="Ecs0758"
complement(6092. .7372)
/note="probable transport protein, similar to
glutamate/aspartate transport proteins (proton glutamate
symport proteins) e.g. [Bacillus stearothermophilus]
gi1124671|sp|P29431|GLYT_BACST percent identity 38 in 416
aa, also similar to C4-dicarboxylate transpor."
/codon_start=1
/transl_table=1
/evidence=not_experimental
/product="putative transport protein"
/protein_id="BAB34181.1"
/db_xref="gi:13360217"
/translacion="MKKISLTMTMLALVLCMIGVNLNNTASBPETAKLYAOEISFT
TIFRLIKMILAPLVSTLVGIAKKGDALKGRIFSTPLFLICASLSTALGLITV
NFRPFGCINFEVAGHGETGVAADEPFLTKYISNAPPSIYDAMAHNEIQIIVFSI
VLSGSLTAIEKGSALVHALDSLAMHMKLKTQVWLPALVYPAISALIERGALVAV
VLSAIEFGEFFYTMLLWLLIGLAIYVYGPICIRLRLTSPALIAFTSSSAAPPE
GTLLEKLEQSPVRIASFVPLIGYSFNVLGSMAGISFATVFIACQCNHLSIGQIIM
LILMLTSSKMGAGVPRASHVVIYAATLGNINPEAGLILMGDPELIDMGKSNATVMSN
AMGAANYSRWEGEFEGCGKGLKPESNAVALP"
complement(7533. .7850)
/gene="Ecs0759"
complement(7533. .7850)
/gene="Ecs0759"
/note="unknown"
/codon_start=1
/transl_table=1
/evidence=not_experimental
/product="hypothetical proteino"
/protein_id="BAB34182.1"

```

RESULT	14	29-SEP-1997
E00893	LOCUS	
E00893	1121 bp	DNA
E00893	Genomic DNA encoding biotin Synthetase.	
E00893	ACCESION	
E00893.1	GI:2169154	
JP 1986149091-A/1.	KEYWORDS	
Escherichia coli.	SOURCE	
Escherichia coli	ORGANISM	
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.		
REFERENCE		
1 (bases 1 to 1121)		
Hirono,Y., Kojima,T. and Kimura,H.		
DUPLEX DNA TO CODE BIOTIN SYNTHASE, BACTERIUM CONTAINING SAME AND		
PRODUCTION OF BIOTIN		
Patent: JP 1986149091-A 1 07-JUL-1986;		
JOURNAL		
NIPPON SODA CO LTD		
OS Escherichia coli		
PN JP 1986149091-A/1		
PD 07-JUL-1986		
PF 24-DEC-1984 JP 1984272605		
PI HIRONO YOSHIIKO KOJIMA TAKAKAZU, KIMURA HITOSHI PC		
C12N15/00,C12N1/20,C12P13/18,(C12N1/20,C12R1.19),(C12P13/18, PC		
C12R1.19);		
CC strandedness: Double;		
CC topology: Linear;		
CC hypothetical: No;		
CC anti-sense: No;		
CC *source: strain=Escherichia coli Nsl01;		
CC feature is identified by experimental;		
PH key	Location/Qualifiers	
FT CDS	42..1079	
FT	/product='biotin synthetase'.	
FEATURES		
source	Location/Qualifiers	
1..1121	/organism='Escherichia coli'	
BASE COUNT	/db_xref='taxon:562'	
ORIGIN	289 a 296 c 325 g 211 t	
Query Match	88.0%; Score 22; DB 10; Length 1121;	
Best Local Similarity	100.0%; Pred. No. 0.1;	
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
OY 4 cagcaggtgcacgcagcatt 25		
Db 123 CACGAGGTGCATGCGCAGCAT 144		
RESULT 15		
MAV250020	2923 bp	DNA
LOCUS	Mycobacterium avium subsp. paratuberculosis	BCT
DEFINITION	1S900, Locus 7.	26-SEP-2000
ACCESSION	AJ250020	
VERSION	AJ250020.1	GI:8919146
KEYWORDS	2599 gene; 2600 gene; insertion element; insertion sequence IS900;	
SOURCE	p43 gene; transposase.	
ORGANISM	Mycobacterium avium subsp. paratuberculosis.	
	Mycobacterium avium subsp. paratuberculosis	
	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;	
	Actinomycetales; Corynebacteriaceae; Mycobacteriaceae;	
	Mycobacterium; Mycobacterium avium complex (MAC).	
REFERENCE	1 (bases 1 to 2923)	
AUTHORS	Bull,T.J., Hermon-Taylor,J., Pavlik,I., El-Zaateri,F. and Tizard,M	
TITLE	Characterization of IS900 loci in mycobacterium avium subsp.	
JOURNAL	Paratuberculosis and development of multiplex PCR typing	
MEDLINE	Microbiology 146 (pt 9), 2185-2197 (2000)	
	20431891	

```

REFERENCE      2 (bases 1 to 2923)
AUTHORS        Bull,T.J.
TITLE          Direct Submission
JOURNAL        Submitted (29-NOV-1999) Bull T.J., Surgery, St.George's Hospital
FEATURES       Medical School, London, SW17 0RE, UNITED KINGDOM
SOURCE         Location/Qualifiers
               1. 2923
               /organism="Mycobacterium avium subsp. paratuberculosis"
               /sub_species="paratuberculosis"
               /db_xref="taxon:1770"
               complement(11..466)
               /gene="2600"
               complement(11..466)
               /function="unknown"
               /codon_start=1
               /transl_table=1
               /product="hypothetical protein"
               /protein_id="CAB96064.1"
               /db_xref="GI:8919147"
               /translation="MYLVEIGTVDLPVKGAVATILYFAVGAVALVGYANDLLT
               PGRLQVLFIDRRPNAVVAGAMTATATVITTAIANSYSQGGGLGVAVYGLMGVI
               LIGVALLAMHLLIPGSHFHEVERPOLHPSFAVALLILAVGGVTAAS"
               473..1922
               /organism="Mycobacterium avium subsp. paratuberculosis"
               /insertion_seq="IS900"
               /sub_species="paratuberculosis"
               /db_xref="taxon:1770"
               686..1906
               /gene="p43"
               686..1906
               /gene="p43"
               /transl_table=1
               /product="transposase"
               /protein_id="CAB96065.1"
               /db_xref="GI:8919148"
               /translation="MTVEVYVAPVWAGVYAGACADHYCWITINDAQRLLSRYVAND
               AALLILAAVTTIADGGVYVNAIDNAGGALLIALIAAGORLLYIPGTVHAAAS
               YVECKTAKDAIITADAKMRHDIOPRAGDDIIVELRITSRSDLVADRTIALEP
               NARPAAGTILSALERAEDYKSKRAALILITGYOTDALRSAGARVAFLKRRARND
               IVAATLQANAHSHIVPGQDLATVVARLAKEMVALDTEIGDDAIEIERPRHRA
               EITLSPGEVYLGAELFATGMAAFASADRLAGVAGIAPVPRDSGRISGMLKRR
               RYDRRLRACVLSALVSIRTPSSRTTYDKRRTGKRHTOAVIALARRRLVLMMLR
               DHAYHPATTTAA"
               complement(11963..2394)
               /gene="2599"
               complement(1963..2394)
               /gene="2599"
               /function="unknown"
               /codon_start=1
               /transl_table=1
               /product="hypothetical protein"
               /protein_id="CAB96066.1"
               /db_xref="GI:8919149"
               /translation="VSRNRLFVIALGALAAVAVCLVTGIILOKNIASVYAGHYEVA
               RDVNGTRYQCTGSEYVADTLESEAEPAAADNKTETRYSDIYIVPDNDNYPCSI
               RVPLISAGYSHRAVFLVPGPFPGSGAGATPGGPGGT"
BASE COUNT    473 a 1005 c 1000 g 445 t

Query Match      80.0%; Score 20; DB 3; Length 2923;
Best Local Similarity 100.0%; Pred. No. 0.97;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      4 cagcaggtgcacgcgcagca 23
        |||||||||||||||||||
Db      125 CAGCAGGTGCATGCCACAGA 144

RESULT 16
PSEHRPIA/c

```

```

LOCUS          PSEHRPIA      3603 bp      DNA      BCT      21-OCT-1993
DEFINITION    Pseudomonas syringae hrp1 genes, complete cds.
ACCESSION     L11582
VERSION       L11582.1
KEYWORDS      hrp1 gene; protein translocation.
SOURCE        Pseudomonas syringae (strain 61) DNA.
ORGANISM      Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
               Pseudomonas.
REFERENCE     1 (bases 1 to 3603)
AUTHORS       Huang,H.-C., Xiao,Y., Lin,R.-H., Lu,Y., Hutcheson,S.W. and
               Collmer,A.
TITLE         Characterization of the Pseudomonas syringae pv. syringae 61 hrp1
               and hrp1 genes: homology of Hrp1 to a superfamily of proteins
               associated with protein translocation
JOURNAL       Mol. Plant Microb Interact. 6, 515-520 (1993)
FEATURES      Location/Qualifiers
SOURCE        1..3603
               /organism="Pseudomonas syringae"
               /db_xref="taxon:317"
               1330..3417
               /gene="hrp1"
               1330..3417
               /gene="hrp1"
               /note="hypersensitive response and pathogenicity gene,
               membrane-spanning protein; hrp1 superfamily proteins are
               associated with protein translocation; putative"
               /codon_start=1
               /transl_table=1
               /product="Hrp1"
               /protein_id="AAA03015.1"
               /db_xref="GI:151268"
               /translation="MNRVINFNLNVALSAMRSELVGFVAVIAIVEMITPLPTGLID
               VLAIVNICISLIMLAMHLPRLAEPSPFPAVILLTMRPLASGSTRILLINQAG
               HIVEAFQGVVGNLAVGLVFLITVYNPLVITKGERARVAGARTTIDAMKQMS
               IDSDELRLNITVHEARRKRAELNKSOLFQMGAMKFFVGDHIASLIIVAINMIGI
               SIGVQHNMAAGDQLQYVTTITGDGLIAQIPALLISVTSGMTITRVPTGAEVANI
               DEORTAQAIPMNKEDLRTFSPRQVLOFHGQDSAQIEALVSEIRRRRNLLVVOY
               GLTLPSEITIEHVDIADDERFVYDVPMKATFTQSHVAVEAROLEGNLPAIIPGN
               TDREDDQWMLPABOSGELINPVSTLIIIRMRALOSCAPQIGIQEIKLISLES
               BOPELAGQMWVLTITRESAVLORLAECVPLRAIVIAETLIEHQHREDNVITDY
               VRIALRSQIYHOYCGAGLOVWLTVPSESLIDGIRQOTETFPALSWETSOMLVQO
               LHAIPVAPQOAVLVAQDLRSRLRLILKEETTHVPLSFATISNAKVKWGRDIL
               EDLLEPLDNEHAA"

```

```

BASE COUNT    772 a 1093 c 1035 g 703 t

Query Match      76.0%; Score 19; DB 3; Length 3603;
Best Local Similarity 100.0%; Pred. No. 3.2;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      6 gcaggtgcacgcgcagcat 24
        |||||||||||||||||||
Db      1522 GCAGGTGCATGCCACGAT 1504

RESULT 17
MTCY180/c
LOCUS          MTCY180      44201 bp      DNA      BCT      17-JUN-1998
DEFINITION    Mycobacterium tuberculosis H37Rv complete genome; segment 85/162.
ACCESSION     Z97193
VERSION       Z97193.1
KEYWORDS      GI:3261816
SOURCE        Mycobacterium tuberculosis.
ORGANISM      Mycobacterium tuberculosis
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Mycobacteriaceae;
Mycobacterium; Mycobacterium tuberculosis complex.
REFERENCE     1 (bases 1 to 44201)
AUTHORS       Cole,S.T., Brosch,R., Parkhill,J., Garnier,T., Churcher,C.,
               Harris,D., Gordon,S.V., Eigemeier,K., Gas,S., Barry III,C.E.,

```

TITLE
 Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence
 JOURNAL
 Nature 393 (6685), 537-544 (1998)
 MEDLINE
 96295987
 REMARK
 Erratum: [[published erratum appears in Nature 1998 Nov 12:396(6707):190]]
 2 (bases 1 to 44201)
 REFERENCE
 Parkhill, J.
 DIRECT SUBMISSION
 Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
 On Jun 27, 1998 this sequence version replaced gi:2225942.

COMMENT
 Notes:
 Details of M. tuberculosis sequencing at the Sanger Centre are available on the World Wide Web.
 (URL, http://www.sanger.ac.uk/Projects/M_tuberculosis/) CDS have been renumbered from the original cosmid submissions but the old gene designations are in brackets after the new gene numbers.
 Gene prediction was based on a Hidden Markov Model of TB genes implemented in TBparse (Krogh) supplemented with visual inspection of positional base preference in codons, especially where there is an increase in the observed/expected third position G + C.
 CAUTION: In some cases we may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, or ttg) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

FEATURES

source
 1..44201
 /organism="Mycobacterium tuberculosis"
 /strain="H37Rv"
 /db_xref="taxon:1773"
 <1..42142
 /organism="Mycobacterium tuberculosis"
 /strain="H37Rv"
 /db_xref="taxon:1773"
 /clone="Y180"
 23..460
 /gene="Rv1873"
 23..460
 /gene="Rv1873"
 /note="Rv1873, (MTCY180.45c), len: 145. Unknown, TBparse score is 0.909"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein Rv1873"
 /protein_id="CA010052.1"
 /db_xref="GI:2225987"
 /db_xref="SPTREMBL:O07756"
 /translation="MKSASDPDLKRFVYAAQPVRSVEELRACRKGHMVFYFQ LKLGSSPLFAVRGISELEAOAYLOHDLIGRLHCTGLVNOYGRSIEELFGPPD LKLGSSMTLFAARTANDQFVALAKYGGGDRRLVALAVT"
 517..520
 /note="possible RBS, GGAG, for Rv1874"
 533..1219
 /gene="Rv1874"
 533..1219
 /gene="Rv1874"
 /note="Rv1874, (MTCY180.44c), len: 228. Unknown, TBparse score is 0.928"
 /codon_start=1
 /transl_table=11

RBS
 /product="hypothetical protein Rv1874"
 /protein_id="CA010067.1"
 /db_xref="GI:2225986"
 /db_xref="SPTREMBL:O07755"
 /translation="MLMRPEDDDCARQAVADALLGICVAGLSTINVRDSTVRL MLITLTPPVAAVVS.LMTQOCYGEQVAAALRLAOCDELCALYTESVPLTFPSIYE GSRPTGLANIALIRPDPIDQATWLTWRQDHQVALEAOATGTYTONVYRALITE APGAGIVVEELFPAVATDLKAFGAADDNDLRNRSIMVASTAFGANONIDIVPTS RYFRTPRPKD"
 1218..1221
 /note="possible RBS, GAGG, for Rv1875"
 1230..1673
 /gene="Rv1875"
 1230..1673
 /gene="Rv1875"
 /note="Rv1875, (MTCY180.43c), len: 147. Some similarity to G953841 MTC165.22 (147 aa) opt: 178 z-score: 240.3 E(): 7.4e-06; 26.9% identity in 130 aa overlap. TBparse score is 0.908"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein Rv1875"
 /protein_id="CA010051.1"
 /db_xref="GI:2225985"
 /db_xref="SPTREMBL:O07754"
 /translation="MTTLNDEAALAAEAERGLAVSTVRADGTQASLVNGLLPHPV GEPGLGTGYGKVLGMLRARPOLAVTFRNGQWATVEARQALVGDPDRPVLVDGER LRLIRVEVFAAGCTHDDMDQDEYDRVMAQEQRAVAVLITPRYSNG"
 2179..2183
 /note="possible RBS, AGGAG, for Rv1876"
 2189..2668
 /gene="bfrA"
 2189..2668
 /gene="bfrA"
 /note="Rv1876, (MTCY180.42c), bfrA, len: 159. Function: bacterioferritin (bfr) (159 aa) opt: 958; E(): 0; 90.6% identity in 159 aa overlap. TBparse score is 0.913"
 /codon_start=1
 /transl_table=11
 /product="bfrA"
 /protein_id="CA010050.1"
 /db_xref="GI:2225984"
 /db_xref="SPTREMBL:O08465"
 /translation="WQSDPDVRLINQRLSELTAIINOYFLHSMQDNMGTELAHT RAESFDEMRHAEIETRIILLDGLPNYQISLIGTQLRQEPADAIEYDVLNLRK PGIVMCREKQDTSVAVLEKIVADEEHHIDYLETQLEMLKGLBELTSACVSRPP"
 2742..2746
 /note="possible RBS, GAGG, for Rv1877"
 2753..4816
 /gene="Rv1877"
 2753..4816
 /gene="Rv1877"
 /note="Rv1877, (MTCY180.41c), len: 687. Integral membrane protein similar to many antibiotic and drug efflux proteins. M. tuberculosis relatives include: MTCY3012.01, MTCY98.0002c and MTCY369.27c. Contains P500217 Sugar transport proteins signature 2 (P500217). FASTA results: 056175 (357 aa) opt: 895; E(): 0; 34.7% identity in 528 aa overlap. TBparse score is 0.916"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein Rv1877"
 /protein_id="CA010049.1"
 /db_xref="GI:2225983"
 /db_xref="SPTREMBL:O07753"
 /translation="MAGPTAPPTAPAIKAGPLSPVRRNITFTALPGVIAATGQ TIIVPALPTIVALELSTVDQSWAVSYLLGTVVAVVAGKGLDLGRNRVLLGSVYF TVGSLVGLSOTMTMLAISRALQGVAGAIISTAVLAALAEVPLDRKRGVGLGAVF GNVNTPGILGMLTDYLSMRNFAFNIVPSIAVLVAATAVPALARPKEVLDYGLI LVIAVATTAIIMATISMGCTTVMGSAFTIYGLIGAAVGAFFVLEGRRAAAILPRL FGSFPAVCCVLSFVYGFANLGLTFPVPIYICGYGASATAGTASGLTRPLVIGLIIAST GTGLVIGRTGRYKIFPVAGMALNAVAFILMSQMDERTPPLDLSLTVLVLAGIGLSMQ

VLVLIWNTSSFEEDGVATSGVTFRRVVGASFGTATFGALFVNFELDRIGSLTSGAV
 EPPAVPSAVLHOLPQSMAPITVRAVSESLTOVLCASVTAVVGGTIALIREVPLTD
 IHDADDDIGDEGVRAESPEDVLEIAVRRLPGVGRVRLIATOPGCGVATALL
 RIYQORLEFGEAVRLLDGRHLHPYGVFEFPEVPEVNOTGTGAARDLITLITPSGHQV
 DSIATVILKRWLLDHLAVAPGLKRPDPHOFALAOHTVADLVORFDMEDIGLSESRQ
 LAATP"

misc_feature
 3122..3199
 /gene="RV1877"
 /note="PS00217 Sugar transport proteins signature 2"

gene
 4871..6223
 /gene="glna3"
 4871..6223
 /gene="glna3"
 /note="RV1878, (MTCY180.40c), len: 450, glna3, similar to
 many glutamine synthetases. FASTA results: GLNA_BACCE
 P19064 (EC 6.3.1.2) (443 aa) opt: 497; E(): 5.2e-23; 29.0%
 identity in 331 aa overlap. Tblast score is 0.995. Also
 similar to C-terminus of FLUG_EMENT P38094 flug protein.
 z-score: 357.9 E(): 6.4e-13 (29.9% identity in 394 aa
 overlap) (note that the downstream ORF MTCY180.39c is
 similar to the N-terminus) and to several other M.
 tuberculosis proteins eg. MTCY427.03c, MTCY427.01 and
 MTCY190.31"
 /codon_start=1
 /transl_table=11
 /product="glna3"
 /protein_id="CAR10048.1"
 /db_xref="GI:2225982"
 /db_xref="SPTREMBL:007752"
 /translation="MTATPLAAAIQAOLEAGSDVDTGVVNPAGLITQAKVPIRRTN
 TEANPLIGASPVWTFPCIDQSIATFADISVVDGRLRLDLSARLTIGDGLAMPACF
 FEODGTPVACSKGTLSRIEALADAGDAVIGHVEFLIDADQQRPLSTLMQYGV
 AGVLEHFAFVQVNAATAGATAEOPHEVCANPEFELSLAPVPAADLVLTLL
 IGTRARHGLRVLSPAPRAGSICSGAHOHPSLTISEGLFSGGGAAGMSAEAAV
 AGVIRGLPDAQGLICGSYISGLRMRPGMAGIYACVNEEAVRFEYKGGAGSAYCG
 NYEVKVVDSANPPLASAIILGLADGKTAVLEPSETVPTOLSDVDRACITLL
 ADQADATIAVIDSKILRCILDDPVVAIVAVROLEHRYGDLDPALADKFRMAWV
 6213..6216
 /gene="glna3"

RBS

Query Match
 Best Local Similarity 76.0%; Score 19; DB 3; Length 44201;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 cgcagcagtgatcgatcgca 20
 |||||
 Db 25769 CGCAGCAGTGATCGCA 25751

RESULT 18
 AC068068/c
 LOCUS AC068068 37500 bp DNA HTG 28-APR-2000
 DEFINITION Leishmania major chromosome 2 clone L5666 strain Friedlin, ***
 ACCESSION AC068068
 VERSION AC068068.1 GI:7658330
 KEYWORDS HTG; HTGS_PHASE1.
 SOURCE Leishmania major.
 ORGANISM Leishmania major.
 Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
 Leishmania.
 1 (bases 1 to 37500)
 MYLER,P.J., SISK,E., Ruiz,J., Cosenza,P., Cruz,A. and Stuart,K.
 Direct Submission
 Submitted (28-APR-2000) Seattle Biomedical Research Institution, 4
 Nickerson Street, Seattle, WA 98109-1651, USA
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 11 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 12249: contig of 12249 bp in length
 * 12250 12296: gap of unknown length
 * 12297 27295: contig of 14999 bp in length
 * 27296 27342: gap of unknown length
 * 27343 27885: contig of 543 bp in length
 * 27886 27932: gap of unknown length
 * 27933 28468: contig of 536 bp in length
 * 28469 28515: gap of unknown length
 * 28516 28658: gap of unknown length
 * 28659 28703: gap of unknown length
 * 28704 29418: gap of 715 bp in length
 * 29419 29464: gap of unknown length
 * 29465 30625: contig of 1161 bp in length
 * 30626 30671: gap of unknown length
 * 30672 31824: contig of 1153 bp in length
 * 31825 31870: gap of unknown length
 * 31871 33509: contig of 1639 bp in length
 * 33510 33555: gap of unknown length
 * 33556 35178: gap of 1623 bp in length
 * 35179 35224: gap of unknown length
 * 35225 37500: contig of 2276 bp in length.
 location/Qualifiers
 1. 37500
 /organism="Leishmania major"
 /strain="Friedlin"
 /db_xref="taxon:5664"
 /clone="L5666"
 /chromosome="2"

BASE COUNT
 6965 a 11364 c 11233 g 7435 t 503 others

ORIGIN

Query Match
 Best Local Similarity 72.0%; Score 18; DB 73; Length 37500;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcgcagcagtgatcgatcg 18
 |||||
 Db 5534 GCGCAGCAGTGATCGC 5517

RESULT 19
 AC010079/c
 LOCUS AC010079 38000 bp DNA HTG 11-SEP-1999
 DEFINITION Leishmania major chromosome 2 clone L4116 strain Friedlin, ***
 ACCESSION AC010079
 VERSION AC010079.1 GI:5870277
 KEYWORDS HTG; HTGS_PHASE2.
 SOURCE Leishmania major.
 ORGANISM Leishmania major.
 Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
 Leishmania.
 1 (bases 1 to 38000)
 MYLER,P.J., SISK,E., Ruiz,J., Cosenza,P., Cruz,A. and Stuart,K.
 Direct Submission
 Submitted (11-SEP-1999) Seattle Biomedical Research Institution, 4
 Nickerson Street, Seattle, WA 98109-1651, USA
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 3 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * been provided by the submitter.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 1 1078: contig of 1078 bp in length
 * 1079 1201: gap of unknown length
 * 1202 22879: contig of 21678 bp in length


```

/transl_table=11
/product="biotin synthase"
/protein_id="AAF94271.1"
/db_xref="GI:9655583"
/translation="MEYRHNWVAEVKALLDKPMDLLFEAQOVHRLHHPNHVOYST
LISRTGACPEDEKYPQSAHYRTDVDERLMEYERVLDAQAKNSGTRPCGGAAM
KNPERDMPILKEMI RGVKDMGLTGMTPDOAOQLAAGLDYINNLPTSPF
YGNITRTYODRLDITLSHVDAQMK ICSGIIIGESTNDRAGLIVEANLPTSPF
VPIIMLVKVGTPLEQVDYEPDFEVLIVARTMPKSAVRLSAGREKNEQOALC
FMAGANSIEFGCKLTTPNPAEDSDMLFKKLGINRQVAKPDEITENELDRVER
VAARPTASDLFEYDAL"
2533.3687
/gene="VC1113"
2533.3687
/gene="VC1113"
/note="similar to GB:J04423 SP:P12998 PID:145426
PID:490225 GB:U00096; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="8-amino-7-oxononanoate synthase"
/protein_id="AAF94272.1"
/db_xref="GI:9655584"
/translation="MTLRKRSRTEDALNREHSRLTRQLDLSRAQGEIIOODRTYL
MESSNDYLGANDAEIVQAMOTGLARFGSGSPWVTGSSNAHELEHCHWIGYE
RAVLFSSGFSANQALLFTLEKEDLLIDKLNHASMELGMLSPATMKRPFKNDPEHL
ROLHEOSNLSLVTEGVFSMDGCAPLAOIATLVKORAWLMDADAGTGVIGEDAGS
CQAAGIHPELIVVTFGKAFGLAGAAVLCDAHVGDTLTPFARHHVYSTAMPQAAHALT
HAARMIOSMSREOLAEIACDEDCRNIPGVATOTPIKMWLGSSSALSASHEL
KQOQIWVASIRPTVPVGSARLRTITTAHTQAIQRUSEQLHWEKDT"
3549.4487
/gene="VC1114"
3549.4487
/gene="VC1114"
/note="similar to GB:J04423 SP:P12999 PID:145427
PID:490221 GB:U00096; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="biotin synthesis protein B10c"
/protein_id="AAF94273.1"
/db_xref="GI:9655585"
/translation="MGECSTHRACROCATFYHNGCAVPSADTPLIGAVKTCDEEG
HMSMTATELCEVLKKSALAOAFKAAHTDHAFFORDVGRILQKMSCLKGLRVL
DIGCTGYFSALIRERGAOVVCAIDISHMLDQAKORCGDEGMYOLADAEOLPFASAC
EDWFSLSLQWCEDSLPLSEIRVLPKHGQAFSLTLDGSLFELDQAMRSVDHHRH
INQITISNOVKIALAQAGCSQHIDLAITVWEAFAFMLMDKIGKGNHVSSTGL
ISRRLAKVEVLAYQSFRRNOGLVPATYQVCGVHR"
4463.5179
/gene="VC1115"
4463.5179
/gene="VC1115"
/note="similar to GB:S68059 SP:P13000 PID:1490223
PID:145428 PID:460901; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="gethiobiotin synthetase"
/protein_id="AAF94274.1"
/db_xref="GI:9655586"
/translation="MFRGYSPPRNAFIAGTDTVDGKTVASKATLOALAHNTATTCY
KPVAAAGDKTEFGYRNSDALHMAKATVDYEVNVAIVLPTSPHIAKHWEVTD
YALLSNLSKRONAEILVVEGAGCWVPSDSCLSTWKQERLPIVLAVGLKGLG
SHALTAEARADGELVCMANRINNGTEHTATIEHLGRISTPRLGELPYPMKAK
ROELGKFIQDLHLEPPTVA"
complement(5261.5470)
/gene="VC1116"
complement(5261.5470)
/gene="VC1116"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"

```

```

/protein_id="AAF94275.1"
/db_xref="GI:9655587"
/translation="MSRNLTTNKGKHNMAEASHYQMAIDLLCCHLGISEQEARQYGLT
EAKETLQOOQIYDTQHSMLGISAER"
5697.6560
/gene="VC1117"
5697.6560
/gene="VC1117"
/note="similar to GB:M58470 SP:P23894 PID:146412 GB:U00096
PID:1736470; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="heat shock protein HspX"
/protein_id="AAF94276.1"
/db_xref="GI:9655588"
/translation="MKRILFLATNLAVVILSVLNTVAVTGMQSGSLGLVMAA
VFGFGAFSLMSKSMALRSYGVVDTTPRNEHMLLETVRQANDGIGMPTVAI
YDAPDMAFATGAKRDSLVAVSTGLNNMTDEEAHLAEVSHLNGDVTMTMQ
GVNTFVFLSRFLANIVASRDSERGSNNMYEVSMLVELFEGFLASFTIMYSR
HREFADGAQALVGKHKMIALERLKKQGESHLEGSMMAFGIGTKRSLSEIMTTHP
LEKRIALARNM"
6770.7198
/gene="VC1118"
6770.7198
/gene="VC1118"
/note="similar to PID:1653608; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="transcriptional regulator, putative"
/protein_id="AAF94277.1"
/db_xref="GI:9655589"
/translation="MSIMDVQNAQFOOLDKSRILIEIYHPDVFEDAAHRIEFG
DALYQFLYHNYQNVHTCTFIHQEYANEGAFIYMTNHLHPKLAKGEQYDVAGVSHL
HPAGKYVTHRDYEDMGEMLEQPLVIGQVIRAIKRRLQG"
7195.7920
/gene="VC1119"
7195.7920
/gene="VC1119"
/note="similar to SP:P25970; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="oxidoreductase, short-chain
dehydrogenase/reductase family"
/protein_id="AAF94278.1"
/db_xref="GI:9655590"
/translation="MTAVFTGATSGIGKOLADYAKOGNOVIACRNSVLSLHTO
YANIFPLAEVTDHPGTKALAOIPCOPELMTINAGDCFYIDGKMDYTLMARVRIN
VLGVAAVTEGIIOPHLSGHRVATVGSIALPRAEAYGASKAAVAYIARTLQDMR
PLGIEVTTIPRGVATPILDRNTEFAMPPIITTYERRAOKETKAGIARVSOLYPPKRTW
LIRLLCALPRAMQGRIVRLKKA"
7931.9211
/gene="VC1120"
7931.9211
/gene="VC1120"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11

```

Query Match 68.0%; Score 17; DB 1; Length 12891;
 best local Similarity 100.0%; Pred. No. 29;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	gcgcagcaggtgcagc 17
Db	1575	GCACGACGAGTGATCG 1591

RESULT 22
 AK024397 4674 bp mRNA PRI 29-SEP-2000
 LOCUS AK024397/c
 DEFINITION Homo sapiens cDNA FLJ14335 fis, clone PLACE4000411, highly similar

to Homo sapiens mRNA; cDNA DKFZ586D0624 (from clone DKFZ586D0624).

ACCESSION AK024397
VERSION AK024397.1 GI:10436775
KEYWORDS oligo cloning; fls (full insert sequence).
SOURCE Homo sapiens placenta cDNA to mRNA, clone_1lb:PLACE4 clone:PLACE400411.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (sites)
Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y., Nishikawa,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H., Wagaetsuma,M., Hosoi,T., Kaku,Y., Kodaira,H., Kondo,H., Sugawara,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K., Ono,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K., Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagahara,K., Masuno,Y., Ninomiya,K. and Iwayanagi,T.
NEDO human cDNA sequencing project
TITLE Unpublished (2000)
JOURNAL 2 (bases 1 to 4674)
REFERENCE Isogai,T. and Otsuki,T.
AUTHORS Direct Submission
JOURNAL Submitted (23-AUG-2000) to the DDBJ/EMBL/Genbank databases. Takao Isogai, Helix Research Institute, Genomics Laboratory, 1532-3 Yana, Katsurazaki, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp, Tel:81-438-52-3951, Fax:81-438-52-3952)
COMMENT NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert construction; Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing and clone selection; Helix Research Institute (supported by Japan Key Technology Center etc.) and Department of Virology, Institute of Medical Science, University of Tokyo.
FEATURES
source
1.4674
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1lb="PLACE400411"
/clone_1lb="PLACE4"
/tissue_type="placenta"
/note="cloning vector: pME18SFL3"
608..1273
/note="unnamed protein product"
/codon_start=1
/protein_id="BAB14908.1"
/db_xref="GI:10436776"
/translation="MLFSKYLGMTVEOPKGEELSAAIKRIYATKASGKKLQKVT
LKSPRIITLNTNLTNLIENSVIRISYCTADKMDKPAYIAOSOHNSLCHAF
CTRKMAQAVTLVAOAFKYPAPFMQVSKREKRDKASQSGDVIGARQDCTPPILKS
LVATGNLIDLETFAPLSTVSNATNTMDEVPKPPALSGSSVYVSGVASRSVILSL
TSG"

BASE COUNT 991 a 1342 c 1239 g 1102 t
ORIGIN

Query Match 64.08; Score 16; DB 89; Length 4674;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 9 ggtcgcgcgcagcat 24
|||||
Db 2446 ggtgcgcgcgcagcat 2431

RESULT 23
AE005685 10290 bp DNA BCT 28-MAR-2001
LOCUS Caulobacter crescentus section 11 of 359 of the complete genome.
DEFINITION AE005685 AE005673
ACCESSION AE005685.1 GI:13421206
VERSION
KEYWORDS
SOURCE Caulobacter crescentus.

ORGANISM Caulobacter crescentus
Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
Caulobacter.
REFERENCE 1 (bases 1 to 10290)
AUTHORS Nierman,W.C., Feldblyum,T.V., Laub,M.T., Paulsen,I.T., Nelson,K.E., Eisen,J., Heidelberg,J.F., Alley,M.R.K., Ohta,N., Maddock,J.R., Potocka,I., Nelson,M.C., Newton,A., Stephens,C., Phadke,N.D., Ely,B., Debroy,R.T., Dodson,R.J., Durkin,A.S., Gwinn,M.L., Haft,D.H., Kolonay,J.F., Smit,J., Craven,M., Kouri,H., Shetty,J., Berry,K., Ullrich,T., Tran,K., Wolf,A., Vamathevan,J., Ermolaeva,M., White,O., Salzberg,S.L., Venter,J.C., Shapiro,L. and Fraser,C.M.
Complete genome sequence of Caulobacter crescentus
Proc. Natl. Acad. Sci. U.S.A. 98 (7), 4136-4141 (2001)
2 (bases 1 to 10290)
JOURNAL Nierman,W.C., Feldblyum,T.V., Paulsen,I.T., Nelson,K.E., Eisen,J., Heidelberg,J.F., Alley,M.R.K., Ohta,N., Maddock,J.R., Potocka,I., Nelson,M.C., Newton,A., Stephens,C., Phadke,N.D., Ely,B., Laub,M.T., Debroy,R.T., Dodson,R.J., Durkin,A.S., Gwinn,M.L., Haft,D.H., Kolonay,J.F., Smit,J., Craven,M., Kouri,H., Shetty,J., Berry,K., Ullrich,T., Tran,K., Wolf,A., Vamathevan,J., Ermolaeva,M., White,O., Salzberg,S.L., Venter,J.C. and Fraser,C.M.
Direct Submission
JOURNAL Submitted (31-JAN-2001) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
TITLE Location/Qualifiers
JOURNAL 1.10290
source
/organism="Caulobacter crescentus"
/db_xref="taxon:69394"
complement(74..1363)
/gene="CC0102"
complement(74..1363)
/gene="CC0102"
/note="identified by match to TIGR protein family HM
TIGR00563"
/codon_start=1
/transl_table=11
/product="RNA methyltransferase Rmb"
/protein_id="AAK22089.1"
/db_xref="GI:13421207"
/translation="MTOELNDGLPAREALTLIDALSRGGLEDASANGFRELPER
ERAPRALAMAVLRHIGLIDRLATRLAKAPDPVYMLLRGAQAFLEVPAPAAVA
QVAGLAPANKTSRPRGLVNAVLRGLDAPADPPSLAPMIVARVAVAFGEFTAR
TILGOTATPATDLSKRGADLAPLYVLAELAEVLPGLRLRRKGVAMGPFEGTQ
WQDAAAVPARLDVQPGASVLDLCAAPGGKTTDLAAGAOYVALDSSARLRKLS
NLTRGLSLAEIVADAAVWDDIRTFDALILAPCGATGTFRRHPVILMAARPDVSL
NAVORILDAADRLKPGARLYVCVSEPEGSEAOVAFARRODVALEPTAAEAP
SAGPESASTPRGTIRLLPHRRDDGDDFFARFRKL"
1426..1704
/gene="CC0103"
1426..1704
/gene="CC0103"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK22090.1"
/db_xref="GI:13421208"
/translation="WIVGGODRSISTROGAMTDTPPEDDAEVEGAAFGKTLTP
AARALEEAAARAALSLPPEGGRGDPPTFRGDMKKGVAVDF"
1738..3489
/gene="CC0104"
1738..3489
/gene="CC0104"
/note="identified by match to TIGR protein family HM
TIGR00066"
/transl_table=11
/product="gamma-glutamyltransferase"
/protein_id="AAK22091.1"
/db_xref="GI:13421209"


```

27.0% identity in 318 aa overlap. Contains Short-chain
dehydrogenases/reductases family signature P500061."
/codon_start=1
/transl_table=11
/product="gca"
/protein_id="CAAL7306.1"
/db_xref="GI:2909577"
/db_xref="SPTREMBL:O53634"
/translation="MKYVITGAGGMMGSHLAEMLLAAGHDVYATVCRPTIDPSDLOFN
GAYDITDKSVSDSIATFERPDAVFIHLAOSYPVAVSWAPVETLTNNVGTAVFEAL
RRVRPHAKITIVAGSAAYGVDPSEVPIRNERELRLPHYGVSKATDMLAYOYKSY
GMHTVARIENGCGPRKVGDLSDFVRRCMTWLEHPEGSAITGVNLTKRITVDRL
NRALMLLDKGEAGADYNGVSIAYCMGVLLKOVIAACKRDIIVRVPDAILRPTDK
IIVGDSKILAIATIGMOOEICLTQTIADMDYRSKESALMV"
6021..6107
/gene="gca"
/notes="P500061 Short-chain dehydrogenases/reductases
family signature"
6024..6029
/gene="gca"
/notes="ASNI site: ATTAAT linking fragments N/V"
6667..7257
/gene="gmha"
6667..7257
/gene="gmha"
/notes="RV0113, (MTV031.07), len: 196, gmha, Probable
phosphopentose isomerase similar to many, e.g. H. pylori
AE005, FASTA scores: gplAE00051HPAE000596_11 Helicobacter
pylori from bases (192 aa) opt: 451 z-score: 573.4 E() :
1.9e-24. 45.1% identity in162 aa overlap"
/codon_start=1
/transl_table=11
/product="gmha"
/protein_id="CAAL7307.1"
/db_xref="GI:2909578"
/db_xref="SPTREMBL:O53635"
/translation="MCTARAEETFEVETIAKPTIILNDRVLLAARAIGRLIAGYRA
GARVPMCGNSAADAOHFAELTGHILIPDPPIGALALNASHLLTAANDVDYTV
LINALESARGDITLFAISTGNSMSVLRKAKTARELGVTVAMTGSGLAEFADF
LINVPSRGTQIESHIVETIAISEHVAHLFAFRQ"
7289..7861
gene
Query Match 64.0% Score 16; DB 3; Length 13350;
Best Local Similarity 100.0% Pred. No. 96;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 4 cagcaggttcacgcgc 19
|||||
Db 2646 CAGCAGTTCATGCC 2631

RESULT 25
MTV026 23740 bp DNA BCT 24-JUN-1999
DEFINITION Mycobacterium tuberculosis H37Rv complete genome; segment 157/162.
ACCESSION AL022076 AL223456
VERSION AL022076.1 GI:3256026
KEYWORDS
SOURCE Mycobacterium tuberculosis.
ORGANISM Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacteriaceae; Mycobacteriaceae;
Mycobacterium; Mycobacterium tuberculosis complex.
1 (bases 1 to 23740)
Cole,S.T., Brosch,R., Parkhill,J., Garnier,T., Churcher,C.,
Harris,D., Gordon,S.V., Eigmeier,K., Gas,S., Barry III,C.E.,
Tekala,F., Badcock,K., Basham,D., Brown,D., Chillingworth,T.,
Connor,R., Davies,R., Devlin,K., Feltham,T., Gentles,S.,
Hamlin,N., Holroyd,S., Hornsby,T., Jagels,K., Krogh,A., McLean,J.,
Moule,S., Murphy,L., Oliver,S., Osborne,J., Quail,M.A.,
Rajandream,M.A., Rogers,J., Rutter,J., Seeger,K., Skelton,S.,
Squires,S., Squires,R., Sulston,J.E., Taylor,K., Whitehead,S. and
Barrell,B.G.

TITLE Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence
JOURNAL Nature 393 (6685), 537-544 (1998)
MEDLINE 98295987
REMARK Erratum: [[published erratum appears in Nature 1998 Nov
12;396(6707):190]]
2 (bases 1 to 23740)
REFERENCE
AUTHORS Parkhill,J.
TITLE Direct Submission
JOURNAL Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium
tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome
Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique
Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
On Jun 26, 1998 this sequence version replaced gi:2961395.
Notes:
Details of M. tuberculosis sequencing at the Sanger Centre are
available on the World Wide Web.
(URL, http://www.sanger.ac.uk/Projects/M.tuberculosis/) CDS have
been renumbered from the original cosmid submissions but the old
gene designations are in brackets after the new gene numbers.
implemented in TParse (Krogh) supplemented with visual inspection
of positional base preference in codons, especially where there is
an increase in the observed/expected third position G + C.
CAUTION: In some cases we may not have predicted the correct
initiation codon. Where possible we choose an initiation codon
(atg, ttg, or tgg) which is preceded by an upstream ribosome
binding site sequence (optimally 5-13bp before the initiation
codon) if this cannot be identified we choose the most upstream
initiation codon.
FEATURES
source Location/Qualifiers
1..23740
/organism="Mycobacterium tuberculosis"
/strain="H37Rv"
/db_xref="taxon:1773"
<1..23740
/notes="fragment designated v026. Does not represent a
physical clone"
60..64
/notes="possible RBS for RV3797"
72..1853
/gene="fadE35"
72..1853
/gene="fadE35"
/notes="RV3797, (MTV026.02), len: 593, fadE35, Probable
acyl Co-A dehydrogenase similar to E. coli AibD protein,
gplL20915|ECOAIBD_3 E. coli (aibD) gene, comp (541 aa);
FASTA scores: opt: 455 z-score: 561.5 E() : 9.2e-24. 31.1%
identity in 514 aa overlap. Also similar to MTC1376_2 M.
tuberculosis cosmid (542 aa), AibD, 31.9% identity in 458
aa overlap"
/codon_start=1
/transl_table=11
/product="fadE35"
/protein_id="CAAL7861.1"
/db_xref="GI:2950416"
/db_xref="SPTREMBL:O53577"
/translation="MPEDLEAVDKLPSTPEKARQYOTENYRGAMGLNMYLNDPTLO
FTMAYLRDELAFAPHLTRIGELTGPVTRMAETDRNPRLERYDRMGHDSRYV
LPESFLOSRAVTEARQVARDAAAGVPSLAFADYDLNLOADTGMACALATGGM
VSLVITAYAPDPVREFVLKNSGEMDEAOLTERAGSDGLALFTTAATRSQDVL
LNGFWFASNCAGFAVVIARPEGAPDSTRGVATFVLRPRDSSRNGVTRIRKDL
GRVSAVSEIEFGEVDAEFLVLSGSPSADGSPGGLTRMELTNRLVGRASRLGNA
RVALVSLCYGVORARFGCALIDRPLMRKLAENVVEVALAWVFGFAANRLOPR
CLPDRIVAPVTKTCRIGITIVASDALETHGNGVETIWPVARIIRDAQVNTIWESPD
NITCLDVRGIEQRTAHETLLARLDAAVSSDDDTRLVSRTRIEDDAITAWTKLD
ROLAARLFPAPQMGVYAGALLTEDAANMERATRGDRALVALVARYRYLADQGPL
RGIDADCEALQRFDEIVAGAFVTAEG"
1980..3311
/gene="rv3798"
/notes="rs1557-3rd copy"
1980..3314
repeat_unit
CDS

```

```

/gene="Rv3798" (MTV026.03), len: 444. Putative
/note="Rv3798, (MTV026.03), len: 444. Putative
transposase identical to IS1557 MTCY373_32,
sp|O10621|Y02X.MCTU HYPOTHEICAL 50.2 KD, PROTEIN CY37
(444 aa), 99.8% identity in 444aa overlap and similar to
many, e.g. gp|U57649|DBU57649_1 Dibenzoifuran-degrading
bacterium (163 aa), FASTA scores: opt: 767 z-score: 326.8
E(): 0, 67.3% identity in 168 aa overlap, or to BPS1001.1
B.paratuberculosis insertion sequenc (406 aa), 24.9% identity
in 402 aa overlap."
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv3798"
/protein_id="CA17862.1"
/db_xref="GI:2950417"
/db_xref="SWISS-PROT:O10621"
/translation="MKNVRLPALLGVKRTVIDEIEFEEDDAGGARVAVRPSA
VLRGRCGRKASWYDRAGLRQMSLDMGVEVTELEAPRVNCPHGPVAVPVA
RHHAGHTFAEDTAVMLAVACSTAVCELMRTGATVAVRWAOTERKIDFANL
RRIGIDELSYKRRHRYLTVDVHDGRLVMAAPGHDKATGLFPDAGAEKRAOITV
SADADWIMADVETERCDAIQADPEFVAVVATEALDVERRANNDARAIARTPKG
RGPGKNAPRGEGREARLRKAGYALMKNPEDLTEROSAKLAMIADTPELYAYLL
KESIRHVSVEGEEKOALDRWISMAORCRIPVFEVLARLRHRAVIALDALDGLSO
GLESNTIKRILTRAFGRFPQALIALMLTLAHRPTLPGRRHNPQISO"
/length=4314
/gene="Rv3798"
/complement(382..4935)
/gene="accda"
/complement(<3382..4935)
/gene="accda"
/note="Rv3798c, (MTV026.04c), accda, len: 517. Probable
propionyl-CoA carboxylase. Beta chain, similar to many,
e.g. sp|P53003|PCCB_SACER PROPIONYL-CoA CARBOXYLASE BETA
CH (546aa), FASTA scores: opt: 1549 z-score: 1841.5 E():
0: 50.7% identity in 519 aa overlap, also similar to
u00012.17 M. leprae cosmid B1308 (549 aa), 49.8% identity
in 514 aa overlap and to MTCY71.20 M. tuberculosis cosmid
(548 aa), 48.9% identity in 515 aa overlap"
/codon_start=1
/transl_table=11
/product="accda"
/protein_id="CA17863.1"
/db_xref="GI:2950418"
/db_xref="SPTREMBL:O53578"
/translation="MTVTEVPLHTTAEKLAERLEKEEGEKAARKRKQIPS
ARARIELYDPGSEMEIGALCRTPDPNMLYSGVVTGGLNGRVGFSHDQVIG
GTVEMGKRVKAVLEMCAWGCPIVIGVNDGCGVMDVAVTSLAMVAEGRHLLSG
LVPOISTILGKAGAVYSPIDTDLVAVADGCMFVTPDVITKDTGVEDYSLDELG
ADHRAVSYGNTHOVSESAAYOVYRDLFLSPNCDDKPPVNPGLPEITGHELD
STYVDSDNMAVDMHEVILRTFEDGDFDLDAQAQGLITGVARVGRITGVANOPM
MSGALIDEASDKARFRTSDAFTPLVAVDTPGGLPVEDEKNGITKRGRTLYAV
VEADVPEVTTTIRKSYGAVAVGSKOLADLFPAPTARIVAGDGAOGLMKRFP
DPNAPENQAIRKSFVENYNLMAIPWIAERGFIDAVIDPHETRLRLKSMHLRDKO
LMHVEGRKHG"
/complement(4932..10133)
/gene="pks13"
/complement(4932..10133)
/gene="pks13"
/note="Rv3800c, (MTV026.05c), len: 1733. Probable
polyketide synthase, similar to many, e.g.
gp|M63676|SERERYA.1 S. erythraea first ORF of eryA gene,
involved in complex polyketide formation in erythromycin
biosynthesis. FASTA scores: opt: 1339 z-score: 1411.0 E():
0, 33.6% identity in 1232 aa overlap or to AF040570.9
cluster (3413 aa), 34.3% identity in 1162 aa overlap.
ALSO highly similar to mycobacterial putative polyketide
synthase: MTCY38.19, u00023.3 M. leprae,
MTCY22G10.2, MTCY19H9.2, u00023.6 M. leprae, contains
P500606 Beta-ketoacyl synthases active site"
/codon_start=1
/transl_table=11
/product="pks13"

```

```

/protein_id="CA17864.1"
/db_xref="GI:2950419"
/db_xref="SPTREMBL:O53579"
/translation="MADVASEQENAPAEERALTYPEKRWLRMWGKANGKAPDSIDE
SVYVWELSSSDAVAMAADIEDLTGVLTVAAVAFHPTLESLATRLTIEBEPDIDL
DAEEMSRGTGPAERVIDIAIVGLSTREPEKNTPTQWALQEDIDTIDLPDRMSEF
LEEPRLARVAGARTRGYLDIKGFDSEFPAVAKTADNIDIDPOORALAEVLEAEH
ARIPASSLRGAVGVYIGSSINDYSLVAFSPVAAHPYATIGSSSTIARVSVYFDE
HPSVTTIDACSSSLVATHQVQALRNCEADVVAVAGALIMPWTGPDIGAVILA
PGRIRKFSADAGYTRSEGGMLVLRKVDARDDGAILIATVAGSVNNIDGSSNGLI
APNODAOADVLRRAKAGIDPRVYDITGEAGTGTTLGDDTLEALRLVGRGRPADR
PALLGAVTNGVHLESAGAAASMAKVVALQHDLPISINPAGSEPTIDDAKLKMI
TTPDMPRYGVALAGSVSPFGCANAHVVREVLPRDVEKEPEPEPKAAAEPAE
APTLAGHMLRPEPENGNIITDASVAEEPEPELPGYTEALRLKEALELAEVATPL
VPLAVSLITSKRAAAAEELADNMSPEGQASSLESTGRSLSRNHRSRVAVLAHDH
DEAIGLRVAVAGKQAPVSVYDGEVTTGTPYVWVLAGRAGORHRRKKSILYLRNVEAM
IKRVALVODELGSVLELIDDADQVIEITTOYTRATQIALDELHNGARPAVY
GOSLEAASAYFAGLSLBDATRALCSHSMGBEALFGEYTRMALVEYSADETR
EYSDPDELEVGVVAAPQVYIGCPPEQVDAILARAEEGKFAKFAFKGSHTSOMD
PLIGETLAELOGIKPTSPCGIFSVHGGRYIKPGEGYIHDEYWKGLRHSYFTIG
IRNANDSGTTTLELARPVALMOVALTTADAGLHDQILPTLARKODEVSSMYTVA
OLYVGHDIITLFSRASGPDDVANTIPTRPKREHLPAPHSGDSTYMPGHNVL
PQGRHVEYAPBDGNDVLAALVRAAAHVLPDQULTAEQNAVPGDGRVLTTRTRP
GASVQVHARIDESFTLVYDALVSAGSESVLPVAVGATVIAVADAAPVAETPAD
ADAEITLSDLTTRVMPSGMTWSPDSGFTIERTGLIVGSMGCEPEDELPVEVLEIL
GIDSLMAVRIRKRVREDPDLPITQITAVRADALVVEKILEYAEVHDEVOQLHEOK
TOTAEIRIAQAOELLHGKGTKEPDESAGVALPSPONGEOPNTPALNDVPPRDA
AERTVTRATVITGKSPGCTFEMELPRIDDEAKATVPRHPRAGSTVYETELGRLPD
EALADVRYVLEAGQIDGQVFTRLRPRPEGKVPVFPVHPRAGSTVYETELGRLPD
TPWYGERVEGSLIEERAQOYVPLKLEMGDGGVYVAVGSLGVATACATLRKGD
VRFVGLIDAVRAGEEIPQTEIRKRMRYVAFARKTNAVITPAIPEQLELDDCO
VRFVLDVAVSOSGOVPAIGIEHQRTSYLDNRADIAQIOQPDGVTLLMADRYDDAI
MPEPRVAVQPGSGMEYVSDLEVPIGEGHIOAIDEPIIAKVGSHMSRALGQIEADR
TSEVGO"

```

Query Match 64.0%; Score 16; DB 3; Length 23740;
 Best Local Similarity 100.0%; Pred. No. 86;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 cgcagcaggtcgcacg 17
 Db 3422 CGACGACAGTGCATCG 3437

RESULT 26
 LOCUS POL245436/c
 DEFINITION Pseudomonas putida OCT plasmid alk genes cluster (alkBGHJLK, alkN
 and alkST genes) and flanking DNA, strain Tfr4-1L (Gpol), formerly
 annotated as Pseudomonas oleovorans.
 ACCESSION AJ245436 J04618 J04619 S50571 X52935 X65936
 VERSION AJ245436.1 GI:5824135
 KEYWORDS acyl-CoA synthetase; alcohol dehydrogenase; aldehyde dehydrogenase;
 alkane-1 monooxygenase; alkB gene; alkBFGHJK regulator; alk gene;
 alkS gene; alkT gene; hypothetical protein; methyl-accepting
 chemotaxis protein; ORF; outer membrane protein; plasmid; recA
 gene; recA gene product; rubredoxin 1; rubredoxin 2; rubredoxin
 reductase; trpA gene; transposase.
 SOURCE Pseudomonas putida.
 ORGANISM Pseudomonas putida
 Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 Pseudomonas.
 REFERENCE 1 (bases 1 to 34902)
 AUTHORS Kok,M., Oldenhuis,R., van der Linden,M.P., Raaljes,P., Kingma,J.,
 van Ielyveld,P.H. and Wiltolt,B.
 TITLE The Pseudomonas oleovorans alkane hydroxylase gene. Sequence and
 expression
 JOURNAL J. Biol. Chem. 264 (10), 5435-5441 (1989)
 MEDLINE 89174581
 REFERENCE 2 (bases 1 to 34902)
 AUTHORS Kok,M., Oldenhuis,R., van der Linden,M.P., Meulenbergh,C.H.,
 Kingma,J. and Wiltolt,B.

```

TITLE
The Pseudomonas oleovorans alkBAC operon encodes two structurally
related rubredoxins and an aldehyde dehydrogenase
JOURNAL
J. Biol. Chem. 264 (10), 5442-5451 (1989)
REFERENCE
89174582
AUTHORS
3 (bases 1 to 34902)
Eggink,G., Engel,H., Vriend,G., Terpstra,P. and Witholt,B.
Rubredoxin reductase of Pseudomonas oleovorans. Structural
relationship to other flavoprotein oxidoreductases based on one NAD
and two FAD fingerprints
JOURNAL
J. Mol. Biol. 212 (1), 135-142 (1990)
MEDLINE
90204534
AUTHORS
4 (bases 1 to 34902)
van Belien,J.B., Eggink,G., Enequist,H., Bos,R. and Witholt,B.
DNA sequence determination and functional characterization of the
OCP-plasmid-encoded alkyl genes of Pseudomonas oleovorans
JOURNAL
Mol. Microbiol. 6 (21), 3121-3136 (1992)
MEDLINE
93086421
AUTHORS
5 (bases 1 to 34902)
Panke,S., Meyer,A., Huber,C.M., Witholt,B. and Wubbolts,M.G.
An alkane-responsive expression system for the production of fine
chemicals
Appl. Environ. Microbiol. 65, 2324-2332 (1999)
JOURNAL
6 (bases 1 to 34902)
van Belien,J.B., Roethlisberger,M. and Witholt,B.
Organization of pseudomonas oleovorans Gpol and Pseudomonas putida
pl genes involved in alkane oxidation
Unpublished
7 (bases 1 to 34902)
van Belien,J.B.
Direct Submission
Submitted (21-JUL-1999) van Belien J.B., Institute of
Biotechnology, ETH, ETH-Hoenggerberg, Zuerich CH-8093, Switzerland
Related sequences: AccNo M83932 (IS53), AccNo AF052751 (IS1384),
AccNo U84154 (IS1491).
FEATURES
source
Location/Qualifiers
join(1..15660,16825..34902)
/organism="Pseudomonas putida"
/strain="Tf4-1L (Gpol)"
/db_xref="taxon:303"
/focus
/note="formerly annotated as Pseudomonas oleovorans Tf4-1L
(Gpol)"
CDS
212..751
/note="ORF1"
/codon_start=1
/transl_table=1
/product="hypothetical protein, 20.4 kD"
/db_xref="GI:5824137"
/translation="MSRYSGEYSAKLQEOAKTOSLITKILIRKAIEDRRLPSL
DALSKRGIPKRSALNMSDADGVISCSYNTSREPYNTSEYSDOLAALAEETNNLPAT
QTLPPKRTTORSQOEISTLKQNDVLTNTLGEVYRAYMQLVARVDEHTNODIRYCO
VLKSHTLAIDRAHETLVK"
748..1671
/note="ORF2"
/codon_start=1
/transl_table=1
/product="hypothetical protein, 35.2 kD"
/protein_id="CAB54044.1"
/db_xref="GI:5824137"
/translation="MIHAKAPRIDLFNMGHTTPPSIPMLFDSRGNECWEINSYLS
IGGANSYGVPAIKYFVSHADLVNFOKYTERKLTLLDTDETLYDFVYKONSRT
ANSTIKTIVRIITALLHEHTQENKFNOLLTTPRAHFOINATEDFYRLNRPPLRF
LTHPCDLRLDPTPTPTPSFINNELNMAHEAIDHYTSNOY1IDRWYTLATLLEHTGSR
IEEAINPASSIINAANGPVNTNIPVKGKYGKGLQVLIPTREYQESINSFSLVP
AIIYASFMTGYSLAKIIDILMRAPLPITP"
2396..2755
/note="ORF3"
/codon_start=1
/transl_table=1
/product="hypothetical protein, 15.3 kD"
/protein_id="CAB54045.1"
misc_feature
misc_feature
CDS
801..1671
/note="GI:5824138"
/translation="MKRPDAKVEKYLTPKPYDPKSIDGLAALVELDIKVAVPDPL
EYFLNKSRRKRAKILIERNGFCLMLKRLSESEKTSPPTEVAIVLTVOELNWLIDGF
DLMRNPFOVLTFRVYA"
2440..4351
/note="IS66 related-like sequence"
2774..4306
/note="ORF4"
/codon_start=1
/transl_table=1
/product="hypothetical protein, 57.8 kD"
/protein_id="CAB54046.1"
/db_xref="GI:5824139"
/translation="MISMPENLPDDPVLLKOLLROMIHERASDCKITVHLEEVALLR
ORLFRKTEONGDADATPOLPLPDEAESIAPLDRADDEVEYAPTRKRRKRPPLADLP
RLEVYELREHETLTCACCGKRNKATIEEYSEDLVLPKQIRVIAKRVKYGCDSEAP
VTADKPAQMIKSMASPSVLAHLTTKYVDGLPLREKVLGRGIDIPROTLARWYI
QSEHFOPLNLMRESLINSRIHODETRVOYLKPEPSSQSMWVQNGPPDRVY
ILFDYATSAQAEVYRLIDGYRGVMDYAGYNAQAODLERLGCMAHARRKVEA
QVQPKGKTGRADILNLINKLYGERDLKSDDEDRKADERSLPLTQIKWVVK
TOPQVYTONALGAKIYLAISMKSLERYVEHGYLPMONNAEERAIREFVIGRKWLES
DTPKCATASQALYSIVETAKANGQEPYAMLHAEIRLPQACSVEDYFALLPMNCTPHI
HS"
4330..4351
/note="identical to right inverted repeat of of IS1131
(IS66 family)"
complement(5034..5612)
/note="ORF5"
/codon_start=1
/transl_table=1
/product="hypothetical protein, 21.8 kD"
/protein_id="CAB54047.1"
/db_xref="GI:5824140"
/translation="MVKTVSLPKPAPDLLHTEIKWPTPIWRRFVPPNITLGIKH
HIVGVMSGDSHLEHEIEAGEKYGIPDSGQGPVPEAKRTLYKALBEKRTFDLY
DEGSMDRIRIKERKLPATAPOLPYCEGANACPPDVGVGFGYEGFLERAMADPKHP
EHDMMVNGYIEFPAPAFDCERVOMPKRKA"
complement(5940..6230)
/note="ORF6: hypothetical 10.7 kD protein; 60% identity to
putative integrase/recombinase Y4rC, Rhizobium sp. NGR224
(fragment)"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="CAB54048.1"
/db_xref="GI:5824141"
/translation="MTLISTPASSRCIAAVWRLISORHRRHATAMHLOAGVDLTIVA
LMIGHESPVTHYVEADLAKERLALAIAPRLRRTYRPTDALLKFLQGL"
complement(6284..7108)
/note="ORF7"
/codon_start=1
/transl_table=1
/transl_except=(pos:complement(7106..7108),aa:Met)
/transl_table=1
/product="hypothetical protein, 31.1 kD"
/protein_id="CAB54049.1"
/db_xref="GI:5824142"
/translation="MISMENMGKIRMYTRDOKLSLHEIAKRTGLARNITRKVAPAE
KOPYVORARLAKNLSPTNATLEQADADSLRKQOQRKALKLAIDIKAGYDGSOL
TATIRWKBQKALQAEVPLSFALGEAPFQWSEBGLIGITIRRMQVANHICASR
AEMLVAVPSQGHMEFDATRAFSALGVPGRGMPDYVARHITRLMFPSPGLIGSAA
PRVYGAEPFGRHRTITGNRAISASRPKAKRLIGMPPRIDVERTGRQPPWFY"
7188..7214
/note="identical to left inverted repeat of IS53"
7266..7353
/note="ORF8"
/translation="Similar to IS1384 internal fragment"
7369..7386
/translation="Alks, positive regulator of alkBFGHKL and
alkST expression"
7424..15611
/evidence="experimental"
/product="alkBFGHKL mRNA"
7520..8725
/genome="alkB"

```

```

CDS
7520..8725
/gene="alkB"
/feature="Integral membrane non-heme iron monooxygenase"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="alkane-1 monooxygenase"
/protein_id="CAB54050.1"
/db_xref="GI:5824143"
/translation="MLEKRVLDAPPEYVDKKYILITIMPATPMIGLMTGM
GIFYGLVLLWGLPLDPMFEDFNNPEVVKLEKERTYRVTITLTPMAYAL
IYAWMGVOTPMWLEIGALSLGIVNGLAINTGELGKRETFDWMATVAVG
YGHFTENKGGHRRDVAIPDPMATSRMGSTYKSTREIPAFIRAMGLEORLSRG
OSVSPDNELLOPMITVILYAVLLAFGPKMLVFLPIOMFGWOLTSANYIEHYG
LROKMGDEYEHOKPHHSMNSNHYISNLVLFILORSHHNAHPTKSYOSLDPCGLA
LPITYGALFLMAMITPMKPSVMDPRVMDAGDLMKIDLDMSMRETYLKKRTSAGH
SSSTSAVAS
8942..9340
/gene="alkF"
8942..9340
/gene="alkF"
/codon_start=1
/transl_table=11
/product="rubeodoxin 1"
/protein_id="CAB54051.1"
/db_xref="GI:5824144"
/translation="MSRYQCPDQYLYDENKGEPEHGHPTSMNDIPRWACPDCAV
Query Match 64.08; Score 16; DB 3; Length 34902;
Best Local Similarity 100.08; Pred. No. 80;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 4 cagcaggtcgcacgc 19
|||||
DB 6204 CAGCAGTGCATCGC 6189

RESULT 27
DMB48C10 49587 bp DNA INV 04-FEB-2000
LOCUS Drosophila melanogaster BAC clone BACH48C10.
DEFINITION
ACCESSION AL133503
VERSION AL133503.1 GI:6594136
KEYWORDS
SOURCE
ORGANISM
fruit fly.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 49587)
Murphy, L., Harris, D., and Barrell, B.
Sequencing the distal X chromosome of Drosophila melanogaster
Unpublished
Sanger Centre, Wellcome Trust Genome Campus, Hinxton Hall, Hinxton,
Cambridge CB10 1SA, U.K.
2 (bases 1 to 49587)
Benos, P.
Direct Submission
Submitted (16-DEC-1999) European Drosophila Genome Sequencing
Consortium
Clone-BACH48C10: Contig ID=1; Length=49587; Status=Finished
Sequence submitted by Takis Benos; EMBL Outstation - The EBI,
Hinxton, Cambridge, CB10 1SD, U.K.
E-mail: benos@ebi.ac.uk on behalf of the European Drosophila Genome
Sequencing Consortium. For further information see the European
Drosophila Genome Sequencing Consortium's web site:
http://edgp.ebi.ac.uk/.
The syntax for the representation of annotation used in this record
is documented at:
ftp://ftp.ebi.ac.uk/pub/databases/edgp/sequence_annotation_README
Coding sequences are predicted from computer analysis, using both
gene and CDS prediction programs and matches to other sequences.
These predictions and matches have been evaluated by the annotators

```

```

FEATURES
source
gene
CDS
and may have been refined by hand (in which case a GeneFINDER
prediction will have no score. The annotators have also used their
judgement on what matches to represent in this record. A far more
complete annotation record is available from FlyBase
(http://flybase.bio.indiana.edu/) through the FlyBase Annotation
Object linked by the db_xref qualifier in the Feature Table.
IMPORTANT: This sequence is NOT necessarily the entire insert of
clone BACH48C10. It may be shorter, since we are minimising the
overlap between clones to 100 bases, by trimming them. Clone
BACH48C10 overlaps to the left with clone 82C7 the true left end of
clone BACH48C10 is unknown
Clone BACH48C10 overlaps to the right with clone BACH7M4 The true
right end of clone BACH48C10 is unknown
Sequence in absolute orientation with respect to chromosome.
location/Qualifiers
1..49587
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BAC BACH48C10"
/complement(530..1394)
/gene="EG:BACH48C10.1"
/complement(join(530..1009,1095..1394))
/gene="EG:BACH48C10.1"
/feature="prediction-(method:""genefinder"",
version:""084"", score:""21.51""");
/prediction-(method:""geneScan"", version:""1.0"",
score:""73.25""")
/codon_start=1
/protein_id="CAB63507.1"
/db_xref="GI:6594143"
/translation="MSKYPNPVTKHFVSEPTKSTCKRCHFDNAGHSHENLMLHLK
RKHDNAVKSVIDEKREIRKALAOQANKRANVPYTKVLASKYPLDPTPKOSFTIEDD
DSOHSKMEPTMTNCEGETGVNDKNNATYSGEYVIGKSEIPEDEYOMKNSAMELK
AVDSVAQITNESSVCYAPSTTARKARARASSAALDOSPDEDFLRYLGSKLG
KYSARNTVQVFNIRLKADMGHEETDSMGDDVYSP"
2911..4557
/gene="EG:BACH48C10.2"
/feature="prediction-(method:""genefinder"",
version:""084"", score:""52.65""");
/prediction-(method:""geneScan"", version:""1.0"",
score:""112.75""");
/match=(desc:""GH19593.5prime", GH
Drosophila melanogaster head port Drosophila melanogaster
cDNA clone GH19593 5prime, mRNA sequence""",
species:""Drosophila melanogaster (fruit fly)"",
ranges:(query:2855..3324, target:EMBL::A138635.1..470,
score:""2296.00"""), (query:3478..3538,
target:EMBL::A138635.453..513, score:""287.00"""),
method:""blastn"", version:""1.4.9""")
/codon_start=1
/protein_id="CAB63508.1"
/db_xref="GI:6594144"
/translation="WTLRFLTPRRIVDYCRRLSDIYALAPPEYOATIRTOGIANP
AEKRYNESILRLCSLESILTPSASCLPDGHLERTIPSGSDSSRYVSRVRYQV
DSRLDPRAVITQILRSCKLPKREKEDQMKRQEMKEKGLVLRKEPVREKRPQV
PVOKRLSALRELFPDARKREHSLIEMRQALARSQASVATLRKIGITAEALVOLLGI
RGTRYDGCNMERYHNCISILKSEDPKVAAGOGALRNFRVFAATYTGISLEGDVL
FSGDSPSCWLAFLSEMRLEHROLRLVPLKALSAVALNFRVRRRYPDEGCARAYL
GSLRVIRAVSEHRLIEDSVDELPAASLODHILAIPTPSDIPKYSOGLFLAPASCSP
DLVEITCNLDVAVAGRSRYRDRLOVERQMLROCOEELGIGRLSKSDSVTPDKMMAI
RHLLDSGVKSCRGSLHTNTYWSVPTGIVCIPDWSRKR"
16590..19135
/gene="EG:BACH48C10.3"
/feature="prediction-(method:""genefinder"",
version:""084"", score:""21.51""");
/prediction-(method:""geneScan"", version:""1.0"",
score:""73.25""")
/codon_start=1
/protein_id="CAB72239.1"
/db_xref="GI:6911910"
/translation="MSRSSSTRECGSDLYDPLAEFLHNVOLYKHTRENIDALNAKFA
NLOEPAMVILIGBSKAEINTTWLGNPTPTSKLFIKITYVYVVCISLMLPLPEYDGL

```

```

TSKHLLEAKRQELMERLNSODDOEDSSLVVERFKRPHYONOTOILQOORLARVHHG
NDLPDSIGSQSSCGCTITRQPKILLRAHLPLMOQRTSEVVISGVALCDALMKALKRQ
LTPDMCEVSTHSGRHIIIPWHTDITGLHVEELFVLLDKPFIIRTHIKQOILRTKFFSL
VEBCGRRLLETFGEYSQCNFRHORCANRPMLOQPPMDSYOLLAEENDQNGVCP
PGRGAVFRNMSRSRSCSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS
VCINIRSVTSEVORSLIMQARPLPHPCPTHNSNSTQASPTSPRTKHNPRKADSDEN
KILLRDAKSSSENNRNIILAEELIGTRGSPFCFVYRAHHGGVAKYTLAVKTFEPA
QLOAFKNEVAMIKKTRHCNITLLFMCVCSPSLAVTQCEGSSSLYKHAVHSETFKLN
TLIDIGROVAGMUTLAKNIIHRDLKSNMIFELHDSVKRIGDFGLAKTRMGEKO
ANOPTGSLMMAPEVIRMOELNPSFQSDYAFGIIVMVELLAECILPYGHISNKQIIF
AVRGRLLPDMSQVSDAPQALRIAEEDICIKTFKPRDLFPLMLNMLNMLRTLPKIH
RSASEPNLTQSOILNDEFLPLSPKTPVFNFNQFFGSGAGNI"
complement(19754..20239)
/gene="EG:BACH48C10.6"
complement(19754..20239)
/gene="EG:BACH48C10.6"
complement(19754..20239)
/gene="EG:BACH48C10.6"
/translation=(method:"genefinder",
version:"084", score:"26.59");
/prediction=(method:"gensecan", version:"1.0",
score:"82.07"); /match=(desc:"SIMILARITY TO RIBOSOMAL
PROTEIN L14", species:"Caenorhabditis elegans",
ranges:(query:20027..20128, target:SPTRMBL::020430:36..3,
score:"94.00"), (query:19952..20011,
target:SPTRMBL::020430:67..48, score:"73.00"),
(query:19835..19924, target:SPTRMBL::020430:109..80,
score:"88.00"), method:"blastx", version:"1.4.9")"
/codon_start=1
/protein_id="CAB63504.1"
/db_xref="GI:6594140"
/translation="MALRLIKTVRQVAOPLAVTLGGQOTOILHTTPACCEIRLRL
RVNDSLGRKMAEGRPRIHVYKRGVIGDKVLAIKCMKGLVGLKQKOK
PKQKPFNSNMLVLDNGSPSLGTRHPIPTILRTILKEKTLAKGADYTKVLAASRY
V"
complement(21238..25103)
/gene="EG:BACH48C10.5"
complement(21238..24581,24649..24951,25046..25103)
/gene="EG:BACH48C10.5"
/translation=(method:"gensecan", version:"1.0",
score:"314.45"); /prediction=(method:"genefinder",
version:"084"); /match=(desc:"GH02860 5prime GH
Drosophila melanogaster head port2 Drosophila melanogaster
cDNA clone GH02860 5prime, mRNA sequence",
species:"Drosophila melanogaster (fruit fly)",
ranges:(query:25045..25290,
target:EMBL::A1064644:369..124, score:"1221.00"),
(query:24786..24952, target:EMBL::A1064644:535..369,
score:"817.00"), method:"blastn", version:"1.4.9");
/match=(desc:"LD12603 5prime LD Drosophila melanogaster
embryo Bluescript Drosophila melanogaster cDNA clone
LD12603 5prime, mRNA sequence", species:"Drosophila
melanogaster (fruit fly)", ranges:(query:25040..25317,
target:EMBL::AA438345:305..28, score:"1363.00"),
(query:24647..24952, target:EMBL::AA438345:604..229,
score:"1422.00"), (query:24549..24579,
target:EMBL::AA438345:635..605, score:"155.00"),
method:"blastn", version:"1.4.9");
/match=(desc:"bs12f02.yl Drosophila melanogaster adult
testis library Drosophila melanogaster cDNA clone bs12f02
5', mRNA sequence", species:"Drosophila melanogaster
(fruit fly)", ranges:(query:23494..24011,
target:EMBL::A1945413:522..5, score:"2254.00"),
method:"blastn", version:"1.4.9");
/match=(desc:"GM04080 5prime GM Drosophila melanogaster
ovary Bluescript Drosophila melanogaster cDNA clone
GM04080 5prime, mRNA sequence", species:"Drosophila
melanogaster (fruit fly)", ranges:(query:22290..22894,
target:EMBL::AA802249:605..1, score:"2971.00"),
method:"blastn", version:"1.4.9");
/match=(desc:"GH25306 5prime GH Drosophila melanogaster
head port2 Drosophila melanogaster cDNA clone GH25306
5prime, mRNA sequence", species:"Drosophila melanogaster
(fruit fly)", ranges:(query:21634..22176,
target:EMBL::A1405337:543..1, score:"2706.00"),
method:"blastn", version:"1.4.9")"
/codon_start=1
/protein_id="CAB72240.1"
/db_xref="GI:6911911"
/translation="MSSMGSPATILNRMGRRKDKSGTGNSSMGKLSKSTOLHHLST
DTDINTSOLEVRITIPVPAATSNAPEDQTFRTYVLLPKDQLFVRLGARVLSKLS
LVCDNKLQADAKYFRSPVDAQVYCESTIGVGLSELRLCHKSSSYNFPDMWK
FORTGVAARLSSSSSDPSSSRHSKRAKRPVSSNSNLSQDMGMYRTTPVYVP
AKVLAAPPKRRKRPAPRPPOVCIIPROSVVDIPRANKSPATAVIYKRPQLCMSPNL
PDLQDCNGNSKSPDESDTGHVATLDLAKPPLAGPEPFRKLQIKKTYAPAP
PDPOSGDNISLALPEVYPTPNILPAPARITPLPIAPVAMPPLAONNGNSPN
GNVSKVLRNTPPEPSPVLSATIEDDNDPTASKQADSGEAPAPSPSPPEEAKSQ
OESSDEDDAIVYVNFKLCOTSKKPEPSAATLAELTALQDIEEDNHVDPAITN
GNATPSSSETSLSNMYLDIPSPDPESQOMQKRNAGLSGSPDLDEVLATITN
OORLTLIRKOPDPQWVEISAKPNMLAFSITTAKTTRIGRADSPRAGASGCTEV
ASSGRALSHRSSSHVSLNKLQNGICIGORSSSELSTIGESSLQSPPEVITLRSK
NSIAESGASAPVYBQRIKQLSLAKQSSVQGVADISSRSEASQOQSTPVPK
SNRGVAKPISPTFRVQEPYPSPALQKSPAVQORSPVPRSLPVAVSVQOAV
SLHQKIVQKSPLESSEKSFVYSKMSPPAVKVPDPPEKVAVEESSGNSSQ
STTPNPPTATVPSPVNLNPNMTVAEIRQSDRDEPAPSSKPLPTTYRSGPSPINFA
method:"blastn", version:"1.4.9")"

```

```

Query Match          64.0%; Score 16; DB 6; Length 49587;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 gtgcattgccagcatt 25
|||||
Db 27002 GTGCATGCCAGCAT 26987

RESULT 29
ATAC009895

method:"blastn", version:"1.4.9")"
/codon_start=1
/protein_id="CAB72240.1"
/db_xref="GI:6911911"
/translation="MSSMGSPATILNRMGRRKDKSGTGNSSMGKLSKSTOLHHLST
DTDINTSOLEVRITIPVPAATSNAPEDQTFRTYVLLPKDQLFVRLGARVLSKLS
LVCDNKLQADAKYFRSPVDAQVYCESTIGVGLSELRLCHKSSSYNFPDMWK
FORTGVAARLSSSSSDPSSSRHSKRAKRPVSSNSNLSQDMGMYRTTPVYVP
AKVLAAPPKRRKRPAPRPPOVCIIPROSVVDIPRANKSPATAVIYKRPQLCMSPNL
PDLQDCNGNSKSPDESDTGHVATLDLAKPPLAGPEPFRKLQIKKTYAPAP
PDPOSGDNISLALPEVYPTPNILPAPARITPLPIAPVAMPPLAONNGNSPN
GNVSKVLRNTPPEPSPVLSATIEDDNDPTASKQADSGEAPAPSPSPPEEAKSQ
OESSDEDDAIVYVNFKLCOTSKKPEPSAATLAELTALQDIEEDNHVDPAITN
GNATPSSSETSLSNMYLDIPSPDPESQOMQKRNAGLSGSPDLDEVLATITN
OORLTLIRKOPDPQWVEISAKPNMLAFSITTAKTTRIGRADSPRAGASGCTEV
ASSGRALSHRSSSHVSLNKLQNGICIGORSSSELSTIGESSLQSPPEVITLRSK
NSIAESGASAPVYBQRIKQLSLAKQSSVQGVADISSRSEASQOQSTPVPK
SNRGVAKPISPTFRVQEPYPSPALQKSPAVQORSPVPRSLPVAVSVQOAV
SLHQKIVQKSPLESSEKSFVYSKMSPPAVKVPDPPEKVAVEESSGNSSQ
STTPNPPTATVPSPVNLNPNMTVAEIRQSDRDEPAPSSKPLPTTYRSGPSPINFA
method:"blastn", version:"1.4.9")"

```


LOCUS ATAC009895 82289 bp DNA PLN 24-JAN-2001
 DEFINITION Arabidopsis thaliana chromosome III BAC T21P5 genomic sequence,
 complete sequence.
 AC009895
 VERSION AC009895.4 GI:12408719
 KEYWORDS HTG.
 SOURCE thale cress.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 Brassicales; Brassicaceae; Arabidopsis.
 REFERENCE 1 (bases 1 to 82289)
 AUTHORS Lin, X., Kaul, S., Town, C.D., Benito, M., Creasy, T.H., Haas, B.,
 Rong, C.M., Koo, H., Fujii, C.Y., Utterback, T.R., Barnstead, M.E.,
 Bowman, C.L., White, O., Nierman, W.C. and Fraser, C.M.
 Arabidopsis thaliana chromosome III BAC T21P5 genomic sequence
 unpublished
 2 (bases 1 to 82289)
 AUTHORS Lin, X. and Kaul, S.
 TITLE Direct Submission
 JOURNAL Submitted (04-SEP-1999) The Institute for Genomic Research, 9712
 Medical Center Dr., Rockville, MD 20850, USA, xlinetlgr.org
 3 (bases 1 to 82289)
 REFERENCE Lin, X.
 AUTHORS Direct Submission
 TITLE Submitted (24-JAN-2001) The Institute for Genomic Research, 9712
 Medical Center Dr., Rockville, MD 20850, USA
 JOURNAL On Jan 24, 2001 this sequence version replaced gi:12280866.
 COMMENT Address all correspondence to:
 Xiaoyang Lin
 The Institute for Genomic Research
 9712 Medical Center Dr.
 Rockville, MD 20850, USA
 e-mail: xlinetlgr.org
 BAC clone T21P5 is from Arabidopsis chromosome III and is near the
 molecular marker ml172.
 The orientation of the sequence is from SP6 to T7 end of the BAC
 clone.
 Genes were identified by a combination of three methods: Gene
 prediction programs including GRAIL (available by anonymous ftp
 from arthur.epm.ornl.gov), Genefinder (Phil Green, University of
 Washington), Genscan (Chris Burge,
 http://genome.stanford.edu/~chris/GENSCAN.html), and NetPlantGene
 (http://www.cbs.dtu.dk/neupgene/cbsneupgene.html), searches of the
 complete sequence against a peptide database and the Arabidopsis
 EST database at TIGR (http://www.tigr.org/tldb/at/at.html).
 Annotated genes are named to indicate the level of evidence for
 their annotation. Genes with similarity to other proteins are named
 after the database hits. Genes without significant peptide
 similarity but with EST similarity are named as 'unknown' proteins.
 Genes without protein or EST similarity, that are predicted by more
 than two gene prediction programs over most of their length are
 annotated as 'hypothetical' proteins. Genes encoding tRNAs are
 predicted by tRNAscan-SE (Sean Eddy,
 http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats are
 identified by RepeatMasker (Arian Smit,
 http://ftp.genome.washington.edu/RM/RepeatMasker.html). Regions of
 genomic sequence that are not annotated as genes but have predicted
 exons by GRAIL are annotated as misc features.
 FEATURES
 location/Qualifiers
 1. 82289
 /organism="Arabidopsis thaliana"
 /db_xref="taxon:3702"
 /chromosome="III"
 /map="ml172"
 /clone="T21P5"
 1. 24217
 /note="Overlap with BAC clone T12T13
 (AC009327:65718..89934)."
 complement(join(<8..215,451..587,675..857,959..1050,
 1459..1593,1682..1748,1837..2055,2146..2340,2492..2879,
 2979..>3100))

gene /gene="T21P5.1"
 complement(<8..>3100)
 /gene="T21P5.1"
 /note="similar to hypothetical protein GB:CA838918
 [Arabidopsis thaliana]"
 complement(join(<8..215,451..587,675..857,959..1050,
 1459..1593,1682..1748,1837..2055,2146..2340,2492..2879,
 2979..3100))
 CDS /gene="T21P5.1"
 /note="unknown protein"
 /codon_start=1
 /protein_id="AAAF01600.1"
 /db_xref="GI:6017117"
 /translation="MSQISGNNNPILSEVYWSLVNKAQKFKIRDLPEYRSRYEN
 YPKPKVYTOQMKROENRQKIVAGIKRMEIGISRIQOLYGYVMRTSDAGYLS
 ESYVEFALITRETRKDFDLNINANKQLFLARFLRMVCLVGRRMVADQPRRS
 LIDCKRFQETDERKEMVVAOELVRLKSDTFRNRPPLSLVLDPNDAQRPAS
 RSLRLTDAITLSSYCNKAVSELTLDSFRMLQCLEWSPSLVSTGAKMGNAPVY
 ARINSQNDPPLPNPKRAVLYRPSITHFLAVLATICEELSHGILLVLSAGKTG
 OTSSPSLSARSATSVENITLRFESHTIKOETESLQITPSSSYIYSDLVPTTRP
 LPTIDSSYVFKNICGAEKGPALLSPSPPLISADFSRQPSGLFTPLTSP
 VQAFCLSTISLSDMETDTEFAEKLSSNMWASTLATDTLHPVSOILKQPFIR
 RLRLRTFCRAVALTTPVFNKNQOPECCSLPSLLPTAPVQSAVFQMANVFAT
 SKETIPQDITMLESP"
 join(<3627..3661,3762..3843,3918..3995,4081..4140,
 4240..4344,4437..4547,4629..4670,4860..4973,5097..5254,
 5361..5494,5576..5682,5785..5827,5903..56144)
 /gene="T21P5.2"
 /note="T21P5.2"
 /gene="T21P5.2"
 /note="predicted by genscan"
 join(3627..3661,3762..3843,3918..3995,4081..4140,
 4240..4344,4437..4547,4629..4670,4860..4973,5097..5254,
 5361..5494,5576..5682,5785..5827,5903..6144)
 /gene="T21P5.2"
 /note="hypothetical protein"
 /codon_start=1
 /protein_id="AAAF01580.1"
 /db_xref="GI:6017097"
 /translation="MDDNRSSSEIRKREIEKPTIASRKLEDTNFKLIQDPPEMALYAK
 VRSEPEIHSLOERIAACIKMKMLIKREYGLGKRCADRVAMIDEKONSYSANEL
 ARRGDULENLIKLAHDKTYEDRPRTFMPSILKILAEYKVPWRNANATISGIIHL
 DQLOMKTKACNDRIRELTSIVENQPTDITISKDNHDPNRKQASIGSDRNDQTN
 EQLLPMEVNTRNPNYHNIIMDTESLFPNNOIGGSGGIPQKRENFGLPSSVAKG
 MIOREKKAENSMFDAYNGNEEFASHVEEGGIDGFIQIDALIGEVEVLCGPVR
 GTTICMOWNRHLEDGROYIEGATPEYIVTADVDKLIAVECICPMDOGROVYRD
 FSGIYSNESVSKDVLINQGLVRLFANDQKIRCGVFNQST"
 6225..6289
 /note="exon predicted by xgrail, quality marginal"
 6384..6503
 /note="exon predicted by xgrail, quality excellent"
 6794..6839
 /note="exon predicted by xgrail, quality excellent"
 6928..7033
 /note="exon predicted by xgrail, quality marginal"
 7293..7411
 /rpt_family="(TAAA)n"
 complement(7488..7547)
 /note="exon predicted by xgrail, quality
 excellent_shadowexon"
 complement(7654..7784)
 /rpt_family="(TAA)n"
 complement(7819..7864)
 /note="exon predicted by xgrail, quality good_shadowexon"
 8054..8113
 /note="exon predicted by xgrail, quality marginal"
 8263..8380
 /rpt_family="(TAAA)n"
 complement(8592..8627)
 /note="exon predicted by xgrail, quality
 marginal_shadowexon"
 <8652..>9527
 /gene="T21P5.3"

```

gene
    <8652..>9527
    /gene="T21P5.3"
    /note="contains zinc finger motif, C3HC4 type (RING
    finger)"
    8652..9527
    /gene="T21P5.3"
    /note="unknown protein"
    /codon_start=1
    /protein_id="AA01602.1"
    /db_xref="GI:6017119"
    /translation="MSGTGNPNMGCTTYSRDCSGQVCVYCPQWCYFIPEPPSPF
    LDDDESSSDSPSLIALIGLAFIVSYTSLSKYCHRRNNSSTAANRI
    SDYTWGGTNNNGAINTNOTIGGAGGDESLIKSTIYKXKMKMPEDSCS
    VCLSEFENESLRLPKCNAPFPCIDIMKSHMNCPCRAFITYSSAVELVDLNO
    OIYENNSISTGDSVVVNLDELNSRKNETVNESGTPFGDAGKRRRAETGVK
    LRRCCGDRYRTGD"
misc_feature
    complement(10838..10896)
    /note="exon predicted by xgrail, quality
    marginal__shadowexon"
    11174..11254
    /note="exon predicted by xgrail, quality good"
    11690..11700
    /note="exon predicted by xgrail, quality marginal"
    13418..13761
    /gene="T21P5.4"
    <11814..>13761
    /gene="T21P5.4"
    /note="similar to unknown protein GB:AA032238 [Arabidopsis
    thaliana]"
    join(11814..12256,12381..12578,12676..13085,13162..13332,
    13418..13761)
    /gene="T21P5.4"
    /note="unknown protein"
    /codon_start=1
    /protein_id="AA01581.1"
    /db_xref="GI:6017098"
    /translation="MAETKKGSESPITIVLVQENRSPDHLGWERELREIDGVN
    KSDQKFNPGTSDLSHNVFEDQSYVDPNPGHSIRDIYEOVCGKPMDSQPPNG
    PAFMSGFQONARKKMSAVANGKFDALPYVKELONFALIDRFASVPGATON
    RPIHISATSHGTNNERKLLIEGPOKTIPELDEAGFTGTCYOCPTLAFPHNLK
    LKYLTFHIDYGLQPKCKEKNLPMYVVBQRMWDLILNPNADPHSDHSECKLVK
    EYELARSPQNNELLFITIDEHGQGVVDVPMPLDQVNDGILGPPPIPEFRNLG
    VVPEFTSPMTEPGTVHSGNSPYLMSQYHSSIPATVKIKFLKDFLTRDSMAGT
    FESVITRNSPDQCEPTLSNVYKMGVIAKENAELSDPELVIIVAAGLDYKNEEL
    LYKICKTCVSDASKTYTKAFDKVEESKARERGGENDIVFCVDDDDHNVKPPP
    SOSPSHATPWSN"
misc_feature
    complement(14488..14563)
    /note="exon predicted by xgrail, quality
    excellent__shadowexon"
    14893..14935
    /note="exon predicted by xgrail, quality
    marginal"
    join(<15141..15386,15675..15872,16258..16667,16780..16950,
    17036..>17427)
Query Match
    Best Local Similarity 100.0%; Pred. No. 68;
    Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 8 agtgcattgcacga 23
Db 30302 AGGTGATGCCGCA 30317
RESULT 30
AC011732 115955 bp DNA HTG 14-MAR-2000
LOCUS Homo sapiens clone RP11-17011, WORKING DRAFT SEQUENCE, 14 unordered
pieces.
AC011732
AC011732.4 GI:7107960
VERSION HTG: HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS human,

```

```

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
    AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
    TITLE Homo sapiens, clone RP11-17011
    JOURNAL Unpublished
    REFERENCE
    AUTHORS 2 (bases 1 to 115955)
    Biren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
    Brown,A., Castle,A., Collins,S., Collins,S., Collins,S., Collins,S.,
    Cooke,P., DeBartolo,K., Dewar,K., Domino,M., Donnell,L., Doyle,M.,
    Ferraire,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D.,
    Galagan,J., Gardyna,S., Grant,G., Haeas,B., Heaford,A., Horton,L.,
    Howland,J.C., Johnson,R., Jones,C., Kani,L., Karas,A., Klein,J.,
    Lechoczy,J., Lien,C., Locke,K., MacDonald,P., Margulis,N.,
    McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J.,
    Morrison,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
    Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
    Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
    Testaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
    Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
    TITLE Direct Submission
    JOURNAL Submitted (13-OCT-1999) Whitehead Institute/MIT Center for Genome
    Research, 320 Charles Street, Cambridge, MA 02141, USA
    COMMENT On Feb 28, 2000 this sequence version replaced gi:6524206.
    All repeats were identified using RepeatMasker:
    Smit,A.F.A. & Green,P. (1996-1997)
    http://ftp.genome.washington.edu/RM/RepeatMasker.html
    ----- Genome Center
    Center: Whitehead Institute/ MIT Center for Genome Research
    Center code: WIBR
    Web site: http://www-seq.wi.mit.edu
    Contact: sequence_submissions@genome.wi.mit.edu
    ----- Project Information
    Center project name: 17.0.11
    ----- Summary Statistics
    Sequencing vector: M13; M77815; 100% of reads
    Chemistry: Dye-terminator Big Dye; 100% of reads
    Assembly program: Phrap; version 0.960731
    Consensus quality: 91165 bases at least Q40
    Consensus quality: 102606 bases at least Q30
    Consensus quality: 109983 bases at least Q20
    Insert size: 87000; agarose-fp
    Quality coverage: 6.0 in Q20 bases; agarose-fp
    Quality coverage: 4.6 in Q20 bases; sum-of-contigs
    -----
    * NOTE: This is a 'working draft' sequence. It currently
    * consists of 14 contigs. The true order of the pieces
    * is not known and their order in this sequence record is
    * arbitrary. Gaps between the contigs are represented as
    * runs of N, but the exact sizes of the gaps are unknown.
    * This record will be updated with the finished sequence
    * as soon as it is available and the accession number will
    * be preserved.
    *
    1 1835: contig of 1835 bp in length
    * 1836 1935: gap of 100 bp
    * 1936 3722: contig of 1787 bp in length
    * 3723 3822: gap of 100 bp
    * 3823 6826: contig of 3004 bp in length
    * 6827 6926: gap of 100 bp
    * 6927 12743: contig of 5817 bp in length
    * 12744 12843: gap of 100 bp
    * 12844 19565: contig of 6722 bp in length
    * 19566 19665: gap of 100 bp
    * 19666 25342: contig of 5677 bp in length
    * 25343 25442: gap of 100 bp
    * 25443 35236: contig of 9794 bp in length
    * 35237 35336: gap of 100 bp
    * 35337 45543: contig of 10207 bp in length
    * 45544 45643: gap of 100 bp

```

```

* 45644 57815: contig of 12172 bp in length
* 57816 57915: gap of 100 bp
* 57916 73012: contig of 15097 bp in length
* 73013 73112: gap of 100 bp
* 73113 87110: contig of 13998 bp in length
* 87111 87210: gap of 100 bp
* 87211 95720: contig of 8510 bp in length
* 95721 95820: gap of 100 bp
* 95821 102140: contig of 6320 bp in length
* 102141 102240: gap of 100 bp
* 102241 115955: contig of 13715 bp in length.
Location/Qualifiers

```

FEATURES

source

```

1. .115955
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="RP11-17011"
/clone_1lib="RPCT-11 Human Male BAC"
misc_feature
1. .1835
/note="assembly_fragment"
1936. .3722
/note="assembly_fragment"
clone_end:T7
vector_side:left"
misc_feature
3823. .6826
/note="assembly_fragment"
6927. .12743
/note="assembly_fragment"
12844. .19565
/note="assembly_fragment"
19666. .25342
/note="assembly_fragment"
25443. .35236
/note="assembly_fragment"
35337. .45543
/note="assembly_fragment"
45644. .57815
/note="assembly_fragment"
57916. .73012
/note="assembly_fragment"
73113. .87110
/note="assembly_fragment"
clone_end:SP6
vector_side:left"
misc_feature
87211. .95720
/note="assembly_fragment"
95821. .102140
/note="assembly_fragment"
102241. .115955
/note="assembly_fragment"
misc_feature
31603 a 25627 c 26285 g 31102 t 1338 others
BASE COUNT
ORIGIN

```

```

Query Match 64.0%: Score 16; DB 62; Length 115955;
Best Local Similarity 100.0%: Pred. No. 64;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Oy 3 gcaagcaggtgcatcgc 18
|||||
Db 106529 GCACGAGTGTCATCGC 106544

```

```

RESULT 31
AC002457/c AC002457 177380 bp DNA PRI 21-DEC-1999
DEFINITION Homo sapiens BAC clone CTB-60P12 from Tq21, complete sequence.
ACCESSION AC002457
VERSION AC002457.1 GI:3947433
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

```

REFERENCE 1 (bases 1 to 177380)
AUTHORS Smith,A., Beck,C. and Gibson,A.
TITLE The sequence of Homo sapiens BAC clone CTB-60P12
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 177380)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (20-AUG-1997) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 3 (bases 1 to 177380)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (02-DEC-1998) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 4 (bases 1 to 177380)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (21-DEC-1999) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Dec 2, 1998 this sequence version replaced gi:2337872.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: saplense@wuston.wustl.edu
----- Summary Statistics
Center project name: H_RG060P12

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GRB/CHR7>, send mailto:egreen@nhgri.nih.gov, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

Clone CTB-60P12 is from the first release of the human BAC library CTB-978SK-B. The library contains cloned DNA from the male fibroblast cell line 978SK. See: Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); U-J. Kim et al., Genomics 34:213-8 (1996). This clone is available from Research Genetics, Inc. (<http://www.resgen.com>).

VECTOR: pBelBAC11

Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is CTB-60N22, 200 bp overlap; the clone sequenced to the right is CTB-137N13, 200 bp overlap. Actual start of this clone is at base position 1 of CTB-60P12; actual end is at 177380 of CTB-60P12.

FEATURES

source

```

1. .177380
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7q21"
/clone="CTB-60P12"

```

```

gene      /clone.lib="CTB-978SK-B"
           complement(131..112890)
CDS        /gene="WUGSC:H_RG060P12.2"
           complement(join(<131..215,1834..1923,31059..31144,
41160..41293,90777..90892,112769..112890))
           /gene="WUGSC:H_RG060P12.2"
           /note="Rapid interacting protein; similar to U73941
(PID:q1916018); H_RG060P12.2"
           /codon_start=1
           /evidence=not_experimental
           /protein_id="AAC82532.1"
           /db_xref="GI:3947434"
           /translation="MASRSLGSLGIRGGGGGKSKLSARNAVERNNLITVCRSV
KVLIDRSCPEPTIDDSPEPNPPALILEQILSHRLKQGVTEGESPEFMDYIRVACR
IVGQNCICSIENNEHNVSSRAKGRAMIRVALMKHLSLEYISTALRDFKTRFRYEDGA
IVGEEANMLAGMLGLNLAIDFSECLKGEGLDGSPFAVIDYTPYLKIONY"
repeat_region 480..791
              /rpt_family="Alu"
repeat_region 4518..4824
              /rpt_family="Alu"
repeat_region 4825..4937
              /rpt_family="(TA)n"
repeat_region 5732..6004
              /rpt_family="L2"
repeat_region 6012..6047
              /rpt_family="L2"
repeat_region 6143..6231
              /rpt_family="L2"
repeat_region 6232..7254
              /rpt_family="MIR"
repeat_region 7272..7412
              /rpt_family="L2"
repeat_region 7420..7813
              /rpt_family="L1"
repeat_region 7814..8150
              /rpt_family="L1"
repeat_region 8151..8265
              /rpt_family="MER1_type"
repeat_region 8294..9183
              /rpt_family="L1"
repeat_region 9314..10441
              /rpt_family="Retroviral"
repeat_region 10437..12545
              /rpt_family="L1"
repeat_region 12928..13352
              /rpt_family="L1"
repeat_region 14327..14443
              /rpt_family="MaLR"
repeat_region 14443..16612
              /rpt_family="L1"
repeat_region 16719..16804
              /rpt_family="L1"
repeat_region 17093..17116
              /rpt_family="AT_rich"
repeat_region 17492..17774
              /rpt_family="AT_rich"
repeat_region 18097..18135
              /rpt_family="Alu"
repeat_region 19255..19594
              /rpt_family="AT_rich"
repeat_region 19900..19946
              /rpt_family="L1"
repeat_region 19966..20245
              /rpt_family="(GA)n"
repeat_region 20549..20614
              /rpt_family="L1"
repeat_region 20684..20816
              /rpt_family="AT_rich"
repeat_region 20817..21123
              /rpt_family="AT_rich"
repeat_region 21124..21134
              /rpt_family="Alu"
repeat_region 21136..21171
              /rpt_family="AT_rich"

```

```

repeat_region 21478..21940
              /rpt_family="(CA)n"
repeat_region 22310..23003
              /rpt_family="L1"
repeat_region 23135..23653
              /rpt_family="L1"
repeat_region 23658..24140
              /rpt_family="MER21_g"
repeat_region 24263..24304
              /rpt_family="L1"
repeat_region 24419..24447
              /rpt_family="AT_rich"
repeat_region 24618..24865
              /rpt_family="AT_rich"
repeat_region 25627..25650
              /rpt_family="Alu"
repeat_region 25906..26152
              /rpt_family="(TTAA)n"
repeat_region 26153..26536
              /rpt_family="L2"
repeat_region 26654..27556
              /rpt_family="AT_rich"
repeat_region 28842..28903
              /rpt_family="Retroviral"
repeat_region 29888..30080
              /rpt_family="(CAT)n"
repeat_region 30382..30669
              /rpt_family="MER1_type"
repeat_region 32790..32960
              /rpt_family="Alu"
repeat_region 33062..33192
              /rpt_family="MIR"
repeat_region 33199..33398
              /rpt_family="L1"
repeat_region 34000..34042
              /rpt_family="L1"
repeat_region 34043..34316
              /rpt_family="(TAAAA)n"
repeat_region 34414..34602
              /rpt_family="Alu"
repeat_region 35064..35263
              /rpt_family="L1"
repeat_region 36255..36571
              /rpt_family="L1"
repeat_region 36571..36571
              /rpt_family="Alu"

```

```

Query Match      64.0%; Score 16; DB 85; Length 177380;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      8 aggtgcatcgccagca 23
Db      48179 AGGTGCATCGCCAGCA 48164

RESULT 32
AC084136/c 180700 bp DNA HTG 14-FEB-2001
LOCUS      AC084136
DEFINITION Homo sapiens clone RP13-506N4, WORKING DRAFT SEQUENCE, 32 unordered
            pieces.
ACCESSION  AC084136
VERSION    AC084136.2  GI:12830241
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 180700)
            Birren,B., Linton,L., Nusbaum,C. and Lander,E.
            TITLE      Homo sapiens clone RP13-506N4
            JOURNAL    Unpublished
            REFERENCE  2 (bases 1 to 180700)
            Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            AUTHORS

```

TITLE
JOURNAL
COMMENT

Anderson, S., Barna, N., Bastien, V., Beda, F., Boguslavsky, L.,
Boukargater, B., Brown, A., Burkett, G., Campopiano, A., Castle, A.,
Choepe, Y., Colangelo, M., Collins, S., Collimore, A., Cooke, P.,
Dearlano, K., Dewar, K., Diaz, J.S., Dodge, S., Ferreira, P.,
Fitzhugh, W., Gage, D., Galagan, J., Gardina, S., Ginde, S., Goyette, M.,
Graham, L., Grand-Pierre, N., Hagos, B., Heaford, A., Horton, L.,
Iliiev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., LaRoque, K.,
Lamarez, R., Landers, T., Lehotzky, J., Levine, R., Lieu, C., Liu, G.,
Macdonald, P., Marquis, N., McCarthy, M., McEwan, P., McKenna, K.,
McPheeters, R., Meldrum, J., Meneus, L., Mihova, T., Mianga, V.,
Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T.,
O'Donnell, P., O'Neill, D., Olivari, T.M., Oliver, J., Peterson, K.,
Pierre, N., Pisani, C., Pollara, V., Raymond, C., Ribbeck, M., Riley, R.,
Rogov, P., Rottman, D., Roy, A., Santos, R., Schauer, S., Severy, P.,
Sougnuez, C., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
Tirelli, A., Travers, M., Trifillo, J., Vassiliev, H., Viel, R., Vo, A.,
Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J.,
Zimmer, A. and Zody, M.

Direct Submission
Submitted (13-OCT-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Feb 14, 2001 this sequence version replaced gi:10800269.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIRB

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center Project name: L11406

Center Clone name: 506_N_4

Summary Statistics

Sequencing vector: plasmid; N/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 167122 bases at least Q40

Consensus quality: 174060 bases at least Q30

Consensus quality: 176531 bases at least Q20

Insert size: 183000; agarose-ftp

Insert size: 177600; sum-of-coverage

Quality coverage: 3.7 in Q20 bases; agarose-ftp

Quality coverage: 3.8 in Q20 bases; sum-of-coverage

NOTE: This is a 'working draft' sequence. It currently
consists of 32 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1 951: contig of 951 bp in length
* 952 1051: gap of 100 bp
* 1052 2064: contig of 1013 bp in length
* 2065 2164: gap of 100 bp
* 2165 3416: contig of 1252 bp in length
* 3417 3516: gap of 100 bp
* 3517 4816: contig of 1300 bp in length
* 4817 4916: gap of 100 bp
* 4917 6297: contig of 1381 bp in length
* 6298 6397: gap of 100 bp
* 6398 9144: contig of 2747 bp in length
* 9145 9244: gap of 100 bp
* 9245 11578: contig of 2334 bp in length
* 11579 11678: gap of 100 bp
* 11679 30905: contig of 19227 bp in length
* 30906 31005: gap of 100 bp
* 31006 32606: contig of 1601 bp in length
* 32607 32706: gap of 100 bp
* 32707 34256: contig of 1550 bp in length

FEATURES
source

* 34257 34356: gap of 100 bp
* 34357 37618: contig of 3262 bp in length
* 37619 37718: gap of 100 bp
* 37719 39131: contig of 1413 bp in length
* 39132 39231: gap of 100 bp
* 39232 41291: contig of 2060 bp in length
* 41292 41391: gap of 100 bp
* 41392 43172: contig of 1781 bp in length
* 43173 43272: gap of 100 bp
* 43273 45866: contig of 2614 bp in length
* 45867 45986: gap of 100 bp
* 45987 49186: contig of 3200 bp in length
* 49187 49286: gap of 100 bp
* 49287 52781: contig of 3495 bp in length
* 52782 52881: gap of 100 bp
* 52882 57288: contig of 4407 bp in length
* 57289 57388: gap of 100 bp
* 57389 62317: contig of 4929 bp in length
* 62318 62417: gap of 100 bp
* 62418 65388: contig of 2971 bp in length
* 65389 65488: gap of 100 bp
* 65489 70061: contig of 4573 bp in length
* 70062 70161: gap of 100 bp
* 70162 75771: contig of 5610 bp in length
* 75772 75871: gap of 100 bp
* 75872 83423: contig of 7552 bp in length
* 83424 83523: gap of 100 bp
* 83524 89716: contig of 6193 bp in length
* 89717 89816: gap of 100 bp
* 89817 97026: contig of 7210 bp in length
* 97027 97126: gap of 100 bp
* 97127 104654: contig of 7528 bp in length
* 104655 104754: gap of 100 bp
* 104755 114340: contig of 9586 bp in length
* 114341 114440: gap of 100 bp
* 114441 126202: contig of 11762 bp in length
* 126203 126302: gap of 100 bp
* 126303 142208: contig of 15906 bp in length
* 142209 142308: gap of 100 bp
* 142309 159777: contig of 17469 bp in length
* 159778 159877: gap of 100 bp
* 159878 178832: contig of 18855 bp in length
* 178833 178932: gap of 100 bp
* 178933 180700: contig of 1768 bp in length.
Location/Qualifiers

FEATURES
source

1. 951
/note="assembly-fragment"
clone_end:SP6
vector_side:left"
misc_feature
1052..2064
/note="assembly-fragment"
misc_feature
2165..3416
/note="assembly-fragment"
misc_feature
3517..4816
/note="assembly-fragment"
misc_feature
4917..6297
/note="assembly-fragment"
misc_feature
6398..9144
/note="assembly-fragment"
misc_feature
9245..11578
/note="assembly-fragment"
misc_feature
11679..30905
/note="assembly-fragment"
misc_feature
31006..32606
/note="assembly-fragment"
misc_feature
32707..34256
/note="assembly-fragment"
misc_feature
34357..37618
/note="assembly-fragment"

```

misc_feature      /note="assembly-fragment"
                  37719..39131
                  /note="assembly-fragment"
misc_feature      39232..41291
                  /note="assembly-fragment"
misc_feature      41392..43172
                  /note="assembly-fragment"
misc_feature      43273..45886
                  /note="assembly-fragment"
misc_feature      45987..49186
                  /note="assembly-fragment"
misc_feature      49287..52781
                  /note="assembly-fragment"
misc_feature      52882..57288
                  /note="assembly-fragment"
misc_feature      57389..62317
                  /note="assembly-fragment"
misc_feature      62418..65388
                  /note="assembly-fragment"
misc_feature      65489..70061
                  /note="assembly-fragment"
misc_feature      70162..75771
                  /note="assembly-fragment"
misc_feature      75872..83423
                  /note="assembly-fragment"
misc_feature      83524..89716
                  /note="assembly-fragment"
misc_feature      89817..97026
                  /note="assembly-fragment"
misc_feature      97127..104654
                  /note="assembly-fragment"
misc_feature      104755..114340
                  /note="assembly-fragment"
misc_feature      114441..126202
                  /note="assembly-fragment"
misc_feature      126303..142208
                  /note="assembly-fragment"
misc_feature      142309..159777

```

Query Match Best Local Similarity 64.0%; Score 16; DB 76; Length 180700;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 agatgcatcgccagca 23

Db 25268 AGTGCATCGCCAGCA 25253

RESULT 33
LOCUS AE003424

DEFINITION Drosophila melanogaster genomic scaffold 142000013386054 section 8

ACCESSION AE003424 AE002566

VERSION AE003424.1 GI:7290316

KEYWORDS HTG.

SOURCE fruit fly.

ORGANISM Drosophila melanogaster

REFERENCE AUTHORS

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 289090)
Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F., George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N., Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X., Burtan,R.C., Rogers,Y.H., Blazer,J.R.G., Chapple,M., Pfeiffer,B.D., Man,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor,Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J., Andrews,P., Plankoch,C., Baldwin,D., Bailly,M., Basu,A., Baxendale,J., Bayraktaroglu,I., Beasley,E.M., Beeson,K.Y., Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkov,D., Botchan,M.R., Bouck,J., Brokstein,P., Brotlier,P., Burtis,K.C.,

TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source

mrna

gene

CDS

Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I., Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M., Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C., Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferriera,S., Fleischmann,M., Foster,C., Gabriellian,A.E., Gary,N.S., Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorell,J.H., Gu,Z., Guan,P., Harris,M., Harris,N.L., Harvey,D., Helman,T.J., Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J., Wei,M.H., Ibegwam,C., Jalali,M., Kalish,F., Karpen,G.H., Ke,Z., Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C., Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A., Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C., McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C., Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L., Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K., Nusskern,D.R., Pacled,J.M., Palazzolo,M., Pittman,G.S., Pan,S., Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K., Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I., Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C., Stapleton,M., Strong,R., Sun,E., Svirska,R., Tector,C., Turner,R., Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wasserman,D.A., Weinstock,G.M., Weissbach,J., Williams,S.M., Woodage,T., Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F., Zaveri,J.S., Zhan,W., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H., Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O., Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
The genome sequence of *Drosophila melanogaster*
Science 287 (5461), 2185-2195 (2000)
20196006
2 (bases 1 to 289090)
Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
Direct Submission
Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA
Location/Qualifiers
1..289090
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="X"
complement(join(<9373..9459,9519..9720,9784..10373,10559..11005,11151..11813,12630..12711,13406..13521,13582..14921,15552..16167,16573..16587))
/gene="EG:BACH48C10.4"
/product="CT33606"
/db_xref="FLYBASE:Fban0014047"
/db_xref="FLYBASE:Fbgn0040390"
complement(<9373..>16587)
/gene="EG:BACH48C10.4"
/note="CG14047"
/map="2F5-2F5"
/db_xref="FLYBASE:Fban0014047"
/db_xref="FLYBASE:Fbgn0040390"
complement(join(9373..9459,9519..9720,9784..10373,10559..11005,11151..11813,12630..12711,13406..13521,13582..14921,15552..16167,16573..16587))
/gene="EG:BACH48C10.4 gene product"
/note="EG:BACH48C10.4 gene product"
/codon_start=1
/db_xref="FLYBASE:Fban0014047"
/db_xref="FLYBASE:Fbgn0040390"
/protein_id="AFA5777.1"
/db_xref="GI:7290317"
/translation="MSTIIAAGVAGAGVVEADGHPQVYFTYLTLEPQORTNS HGHHVATGATATRTADRSNGHNGHGGHVEDEIILLDKRILNKTGTPSSALHL ANMKGLQDQOQDARLAARSOENVASAEQILPNSSEDERITWTROOLEKRT KELLNAKENGSLARPFGKRLSGVEELMSATLSDIYSEAEVTSYIAKATDSDVG AVMNSSRSAGEGNPVITGSRRLHLDPPVSVSLPNHGCSAGVGGSPKPLPPTRRV VAOQCCPPPRKANOOVKATVITTERATIRSSPTGSGSGATGSGVSPAKYAA CTRCPEDPETHAOLSPDKHOTKTLKSPOTTTTLKQDOPDLISMLIGLADIA PATTICQGEITPREENSTPIYAVPTPSPVILTKTTHYWDNSGCSNGGCTISTSTA TTTTKPRAKLIENIPDSDCSPDEEPYRPMSALRRFTMSLEKLPISDDRMRD ADELIDLDLSEFKPNHGDNSPONEDEDEDEADSERALLVNDLGASSSGVIVNGEGLAS

/strain="PEST"
/db.xref="taxon:7165"
/clone="13N18"
/clone_lib="Notredame1"
/note="end : 77"

BASE COUNT 262 a 202 c 192 g 216 t 21 others

ORIGIN

Query Match 60.0%; Score 15; DB 53; Length 893;
Best Local Similarity 100.0%; Pred. No. 5.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 gcacagcagtgatc 17
|||||
Db 709 GCACGAGGTGCATCG 723

RESULT 35
CNS062HL 978 bp DNA STS 11-JAN-2001
LOCUS T3 end of clone XA70AA002F03 of library XA70AA from strain CBS 6340
DEFINITION of Kluyveromyces thermotolerans, sequence lagged site.
ACCESSION AF219383
VERSION AF219383.1 GI:12205529
KEYWORDS STS.
SOURCE Kluyveromyces thermotolerans.
ORGANISM Kluyveromyces thermotolerans.
REFERENCE Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
AUTHORS Malpertuy, A., Florente, B., Blandin, G., Artiguenave, F., Wincker, P.
1 (bases 1 to 978)
TITLE Genomic Exploration of the Hemiascomycetous Yeasts: 10.
JOURNAL Kluyveromyces thermotolerans
PUBMED FBBS Lett. 487 (1), 61-65 (2000)
REFERENCE 1152885
AUTHORS 2 (bases 1 to 978)
Souchet, J.L., Aigle, M., Artiguenave, F., Blandin, G.,
Boloitin-Fukuhara, M., Bon, E., Brothier, P., Casaregola, S.,
de-Montigny, J., Dujon, B., Durrens, P., Lepingle, A., Florente, B.,
Malpertuy, A., Neugeglise, C., Ozier-Kalogeropoulos, O., Potter, S.,
Saurin, W., Tekala, F., Toffano-Nioche, C., Wesolowski-Louvel, M.,
Wincker, P. and Weissenbach, J.
TITLE Genomic Exploration of the Hemiascomycetous Yeasts: 1. A set of
JOURNAL yeast species for molecular evolution studies(1)
PUBMED FBBS Lett. 487 (1), 3-12 (2000)
REFERENCE 1152876
AUTHORS 3 (bases 1 to 978)
Genoscope.
TITLE Direct Submission
JOURNAL Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
segref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT This STS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces marxianus var. marxianus, Kluyveromyces
laetis var. laetis, Kluyveromyces fragilis var. fragilis, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sobriophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.

FEATURES
source
1. .978
Location/Qualifiers
/organism="Kluyveromyces thermotolerans"
/strain="CBS 6340"
/db.xref="taxon:4916"
/clone="XA70AA002F03"
/clone_lib="XA70AA"
/note="end : 73"

STS

BASE COUNT 274 a 192 c 189 g 310 t 13 others

ORIGIN

Query Match 60.0%; Score 15; DB 53; Length 978;
Best Local Similarity 100.0%; Pred. No. 5.3e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 ccacagcagtgatc 16
|||||
Db 753 CCACGAGGTGCATCG 767

RESULT 36
AF219383/c 3069 bp mRNA INV 05-JAN-2000
LOCUS AF219383
DEFINITION Drosophila melanogaster tipop protein mRNA, partial cds.
ACCESSION AF219383
VERSION AF219383.1 GI:6671378
KEYWORDS fruit fly.
SOURCE Drosophila melanogaster
ORGANISM Drosophila melanogaster
REFERENCE Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
AUTHORS Fasano, L., Nhouyvanisvong, M., Caudit, X., Kerridge, S., Jacq, B. and
Vola, C.
1 (bases 1 to 3069)
TITLE leashirt-related paralogs
JOURNAL Unpublished
PUBMED 2 (bases 1 to 3069)
REFERENCE Fasano, L., Nhouyvanisvong, M., Caudit, X., Kerridge, S., Jacq, B. and
Vola, C.
1 (bases 1 to 3069)
TITLE Direct Submission
JOURNAL Submitted (28-DEC-1999) IGPD, CNRS Case 907, Parc Scientifique de
Luminy, Marseille cedex 9 13288, France
REFERENCE 1. 3069
Location/Qualifiers
1. 3069
/organism="Drosophila melanogaster"
/db.xref="taxon:7227"
/chromosome="2L"
/map="40B-40C"
1. .3069
/note="3' end of the gene; similar to leashirt: genomic
DNA deposited under Genbank Accession Number AC006415;
expressed in the Malpighian tubules and stomatogastric
nervous system"

FEATURES
source
1. 3069
Location/Qualifiers
1. 3069
/organism="Drosophila melanogaster"
/db.xref="taxon:7227"
/chromosome="2L"
/map="40B-40C"
1. .3069
/note="3' end of the gene; similar to leashirt: genomic
DNA deposited under Genbank Accession Number AC006415;
expressed in the Malpighian tubules and stomatogastric
nervous system"

CDS
1. 3069
/organism="Drosophila melanogaster"
/db.xref="taxon:7227"
/chromosome="2L"
/map="40B-40C"
1. .3069
/note="3' end of the gene; similar to leashirt: genomic
DNA deposited under Genbank Accession Number AC006415;
expressed in the Malpighian tubules and stomatogastric
nervous system"

BASE COUNT 761 a 917 c 803 g 588 t

ORIGIN

Query Match 60.0%; Score 15; DB 5; Length 3069;
 Best Local Similarity 100.0%; Pred. No. 4.3e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 agcaggtcgcgcgc 19
 |||||
 Db 2114 AGCAGGTGCATCGCC 2100

RESULT 37
 AF274670/c 3369 bp DNA PLN 13-AUG-2000
 LOCUS AF274670
 DEFINITION Tetracentron sinense 26S ribosomal RNA gene, complete sequence.
 ACCESSION AF274670
 VERSION AF274670.1 GI:9799468
 KEYWORDS
 SOURCE Tetracentron sinense.
 ORGANISM Tetracentron sinense.
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 Magnoliophyta; eudicotyledons; Trochodendraceae; Tetracentroideae;
 Tetracentron.

REFERENCE 1 (bases 1 to 3369)
 Fishbein,M., Hibsch-Jetter,C., Soltis,D.E. and Hufford,L.
 Phylogeny of Saxifragales (Angiosperms, Eudicots): Analysis of a
 Rapid, Ancient Radiation
 Unpublished
 2 (bases 1 to 3369)
 Fishbein,M., Hibsch-Jetter,C., Soltis,D.E. and Hufford,L.
 Direct Submission
 Submitted (30-MAY-2000) School of Biological Sciences, Washington
 State University, Pullman, WA 99164-4236, USA

FEATURES
 source 1..3369
 /organism="Tetracentron sinense"
 /specimen_voucher="Qlu 94166 (IND)"
 /db_xref="taxon:13715"
 rRNA 1..3369
 /product="26S ribosomal RNA"
 BASE COUNT 806 a 824 c 1049 g 689 t 1 others
 ORIGIN

Query Match 60.0%; Score 15; DB 13; Length 3369;
 Best Local Similarity 100.0%; Pred. No. 4.2e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 aggtgcgcgcgcgc 22
 |||||
 Db 441 AGGTGCATCGCCAGC 427

RESULT 38
 AF274671/c 3369 bp DNA PLN 13-AUG-2000
 LOCUS AF274671
 DEFINITION Trochodendron aralioides 26S ribosomal RNA gene, complete sequence.
 ACCESSION AF274671
 VERSION AF274671.1 GI:9799469
 KEYWORDS
 SOURCE Trochodendron aralioides.
 ORGANISM Trochodendron aralioides.
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 Magnoliophyta; eudicotyledons; Trochodendraceae; Trochodendroideae;
 Trochodendron.

REFERENCE 1 (bases 1 to 3369)
 Fishbein,M., Hibsch-Jetter,C., Soltis,D.E. and Hufford,L.
 Phylogeny of Saxifragales (Angiosperms, Eudicots): Analysis of a
 Rapid, Ancient Radiation
 Unpublished
 2 (bases 1 to 3369)
 Fishbein,M., Hibsch-Jetter,C., Soltis,D.E. and Hufford,L.
 Direct Submission
 Submitted (30-MAY-2000) School of Biological Sciences, Washington
 State University, Pullman, WA 99164-4236, USA

FEATURES
 source Location/Qualifiers
 1..3369
 /organism="Trochodendron aralioides"
 /specimen_voucher="Qlu 94157 (IND)"
 /db_xref="taxon:4407"
 rRNA 1..3369
 /product="26S ribosomal RNA"
 BASE COUNT 801 a 816 c 1060 g 690 t 2 others
 ORIGIN

Query Match 60.0%; Score 15; DB 13; Length 3369;
 Best Local Similarity 100.0%; Pred. No. 4.2e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 aggtgcgcgcgcgc 22
 |||||
 Db 441 AGGTGCATCGCCAGC 427

RESULT 39
 MTCY9C4 15916 bp DNA BCT 17-JUN-1998
 LOCUS MTCY9C4
 DEFINITION Mycobacterium tuberculosis H37Rv complete genome; segment 113/162.
 ACCESSION Z77250 AL123456
 VERSION Z77250.1 GI:3261617
 KEYWORDS
 SOURCE Mycobacterium tuberculosis.
 ORGANISM Mycobacterium tuberculosis
 Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 Actinomycetales; Corynebacteriaceae; Mycobacteriaceae;
 Mycobacterium; Mycobacterium tuberculosis complex.

REFERENCE 1 (bases 1 to 15916)
 Cole,S.T., Brosch,R., Parkhill,J., Garnier,T., Churcher,C.,
 Harris,S.D., Gordon,S.V., Eiglmeier,K., Gas,S., Barry III,C.E.,
 Tekala,F., Badcock,K., Basham,D., Brown,D., Chillingworth,T.,
 Connor,R., Davies,R., Devlin,K., Feltwell,T., Gentles,S.,
 Hamlin,N., Holroyd,S., Hornsby,T., Jagels,K., Krogh,A., McLean,J.,
 Moule,S., Murphy,L., Oliver,S., Osborne,J., Quail,M.A.,
 Rajandream,M.A., Rogers,J., Rutter,S., Seeger,K., Skelton,S.,
 Squares,S., Squares,R., Sulston,J.E., Taylor,K., Whitehead,S. and
 Barrrell,B.G.
 Deciphering the biology of Mycobacterium tuberculosis from the
 complete genome sequence
 Nature 393 (6685), 537-544 (1998)
 98295987
 Erratum: [[published erratum appears in Nature 1998 Nov
 12:396(6707):190]]
 2 (bases 1 to 15916)
 Parkhill,J.
 Direct Submission
 Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium
 tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome
 Trust Genome Campus, Hinxton, Cambridge CB10 1SA unite de Genetique
 Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
 75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
 On Jun 27, 1998 this sequence version replaced gi:1460072.

Notes:
 Details of M. tuberculosis sequencing at the Sanger Centre are
 available on the World Wide Web.
 (URL, http://www.sanger.ac.uk/Projects/M_tuberculosis/)
 CBS have
 been renumbered from the original cosmid submissions but the old
 gene designations are in brackets after the new gene numbers.
 Gene prediction was based on a Hidden Markov Model of TB genes
 implemented in TParse (Krogh) supplemented with visual inspection
 of positional base preference in codons, especially where there is
 an increase in the observed/expected third position G + C.
 CAUTION: In some cases we may not have predicted the correct
 initiation codon. Where possible we choose an initiation codon
 (atg, gtg, or ttg) which is preceded by an upstream ribosome
 binding site sequence (optimally 5-13bp before the initiation
 codon). If this cannot be identified we choose the most upstream
 initiation codon.

COMMENT

```

FEATURES
source      Location/Qualifiers
            1..15916
             /organism="Mycobacterium tuberculosis"
             /strain="H37Rv"
             /db_xref="taxon:1773"
             <1..13876
              /organism="Mycobacterium tuberculosis"
              /strain="H37Rv"
              /db_xref="taxon:1773"
              /clone="Y9C4"
              <1..40
               /organism="Mycobacterium tuberculosis"
               /strain="H37Rv"
               /db_xref="taxon:1773"
               /clone="Y9C4"
               <1..432
                /gene="Rv2556c"
                /complement(43..432)
                /note="Rv2556c, (MTCY09C4.12), len: 129. Conserved hypothetical protein similar to A6000961|AE000961.7 Archaeoglobus fulgidus (126 aa) fasta scores: opt: 236 z-score: 312.5 E(): 7e-10; 35.8% identity in 123 aa overlap"
                /codon_start=1
                /transl_table=1
                /product="hypothetical protein Rv2556c"
                /protein_id="CAB01048.1"
                /db_xref="GI:1460084"
                /db_xref="SWISS-PROT:O50742"
                /translation="MLDVTARRRIVDTTPAVRAFCIAHDGICNVFPHATGVAII ETGSGDEDLYDTLVRLPPDDRRRAHAGSGHGADHLPAFVAPSVTVFVSGGPUL GTWOSILVLDINDPNRRSVRLSFVES"
                539..1213
                 /gene="Rv2557"
                 /complement(539..1213)
                 /note="Rv2557, (MTCY9C4.11c), unknown, len: 224, highly similar to MTCY9C4.10c (69.2% identity in 224 aa overlap)"
                 /codon_start=1
                 /transl_table=1
                 /product="hypothetical protein Rv2557"
                 /protein_id="CAB01047.1"
                 /db_xref="GI:1460083"
                 /db_xref="SWISS-PROT:O50741"
                 /translation="MTGAGATGALPRMKEGMIVYARSTTQAGSECIDGIANVRDV MPALQMDGICIGSLYDROSRCIATSAKETAEAMHARSREQVTPIDRCALMRFQGP AVEHEWTAAMHRHSABGACVRAVWKPADVDGIEYKSSVLPQIEGLDGFCSA SLVDRITSGRAVSSATFSPDAMERNRDNALKATSLREAGGEILDECEFIATLAL RPELV"
                 1298..2008
                  /gene="Rv2558"
                  /complement(1298..2008)
                  /note="Rv2558, (MTCY9C4.10c), len: 236, highly similar to MTCY9C4.11c (69.2% identity in 224 aa overlap)"
                  /codon_start=1
                  /transl_table=1
                  /product="hypothetical protein Rv2558"
                  /protein_id="CAB01046.1"
                  /db_xref="GI:1460082"
                  /db_xref="SWISS-PROT:O50740"
                  /translation="MPGAGMKRVFGGTGATGALPRHNGSIVYARSTTIEOPLSV DIGIAHRVDVMPALOEIDGCVSLYDROSRCIATSAKETAEAMHARSREQVTPIDRCALMRFQGP AVEHEWTAAMHRHSABGACVRAVWKPADVDGIEYKSSVLPQIEGLDGFCSA SLVDRITSGRAVSSATFSPDAMERNRDNALKATSLREAGGEILDECEFIATLAL RPELV"
                  1298..2008
                   /gene="Rv2559c"
                   /complement(1298..2008)
                   /note="Rv2559c, (MTCY9C4.09), len: 452. Similar to eycAJ_HAEIN P45262 hypothetical protein h11590. (446 aa)

gene
CDS
            1..15916
             /organism="Mycobacterium tuberculosis"
             /strain="H37Rv"
             /db_xref="taxon:1773"
             <1..13876
              /organism="Mycobacterium tuberculosis"
              /strain="H37Rv"
              /db_xref="taxon:1773"
              /clone="Y9C4"
              <1..40
               /organism="Mycobacterium tuberculosis"
               /strain="H37Rv"
               /db_xref="taxon:1773"
               /clone="Y9C4"
               <1..432
                /gene="Rv2556c"
                /complement(43..432)
                /note="Rv2556c, (MTCY09C4.12), len: 129. Conserved hypothetical protein similar to A6000961|AE000961.7 Archaeoglobus fulgidus (126 aa) fasta scores: opt: 236 z-score: 312.5 E(): 7e-10; 35.8% identity in 123 aa overlap"
                /codon_start=1
                /transl_table=1
                /product="hypothetical protein Rv2556c"
                /protein_id="CAB01048.1"
                /db_xref="GI:1460084"
                /db_xref="SWISS-PROT:O50742"
                /translation="MLDVTARRRIVDTTPAVRAFCIAHDGICNVFPHATGVAII ETGSGDEDLYDTLVRLPPDDRRRAHAGSGHGADHLPAFVAPSVTVFVSGGPUL GTWOSILVLDINDPNRRSVRLSFVES"
                539..1213
                 /gene="Rv2557"
                 /complement(539..1213)
                 /note="Rv2557, (MTCY9C4.11c), unknown, len: 224, highly similar to MTCY9C4.10c (69.2% identity in 224 aa overlap)"
                 /codon_start=1
                 /transl_table=1
                 /product="hypothetical protein Rv2557"
                 /protein_id="CAB01047.1"
                 /db_xref="GI:1460083"
                 /db_xref="SWISS-PROT:O50741"
                 /translation="MTGAGATGALPRMKEGMIVYARSTTQAGSECIDGIANVRDV MPALQMDGICIGSLYDROSRCIATSAKETAEAMHARSREQVTPIDRCALMRFQGP AVEHEWTAAMHRHSABGACVRAVWKPADVDGIEYKSSVLPQIEGLDGFCSA SLVDRITSGRAVSSATFSPDAMERNRDNALKATSLREAGGEILDECEFIATLAL RPELV"
                 1298..2008
                  /gene="Rv2558"
                  /complement(1298..2008)
                  /note="Rv2558, (MTCY9C4.10c), len: 236, highly similar to MTCY9C4.11c (69.2% identity in 224 aa overlap)"
                  /codon_start=1
                  /transl_table=1
                  /product="hypothetical protein Rv2558"
                  /protein_id="CAB01046.1"
                  /db_xref="GI:1460082"
                  /db_xref="SWISS-PROT:O50740"
                  /translation="MPGAGMKRVFGGTGATGALPRHNGSIVYARSTTIEOPLSV DIGIAHRVDVMPALOEIDGCVSLYDROSRCIATSAKETAEAMHARSREQVTPIDRCALMRFQGP AVEHEWTAAMHRHSABGACVRAVWKPADVDGIEYKSSVLPQIEGLDGFCSA SLVDRITSGRAVSSATFSPDAMERNRDNALKATSLREAGGEILDECEFIATLAL RPELV"
                  1298..2008
                   /gene="Rv2559c"
                   /complement(1298..2008)
                   /note="Rv2559c, (MTCY9C4.09), len: 452. Similar to eycAJ_HAEIN P45262 hypothetical protein h11590. (446 aa)

misc_feature
            complement(3160..3183)
             /gene="Rv2559c"
             /note="P500017 ATP/GTP-binding site motif A"
             3542..4519
              /gene="Rv2560"
              /complement(3542..4519)
              /note="Rv2560, (MTCY9C4.08c), len: 325. Unknown membrane protein"
              /codon_start=1
              /transl_table=1
              /product="hypothetical protein Rv2560"
              /protein_id="CAB01044.1"
              /db_xref="GI:1460080"
              /db_xref="SWISS-PROT:O50738"
              /translation="MSQPEHGNPNADQGGNAGSYPPPGAPPPPGPPGPT YLPGVNAPPPPGYPPGPPPTPHLQSGSVSDATSMGMRNRTONAVTLVVP VLAVVALAAVAGTAGLVVALSDRATTAATNTSGVSESDITMTPTAGVMTLGYI ALFAVLVIMHAGTILGCLDIADGKVTATFPPRNGLVYTGILLIYAVTFGTGLC VTFGLFGVAFVAFVADRSTSPIDSVKASITFVSGVSLSMIAQLTAVLVGE LICEVGMIGIPEVAALIHVYTRKLSGQVVEAVRAPVGPMPOLA"
              4876..5169
               /gene="Rv2561"
               /complement(4876..5169)
               /note="Rv2561, (MTCY9C4.07c), len: 97. Unknown"
               /codon_start=1
               /transl_table=1
               /product="hypothetical protein Rv2561"
               /protein_id="CAB01043.1"
               /db_xref="GI:1460079"
               /db_xref="SWISS-PROT:O50737"
               /translation="MGIORAVLLIADIGYNNYHMMNRKHLAHOVTLAOLLEVIDA AKGMKLAKLEGMAFFMAPGGOHOCFGMRPRADAPVPHARADOKRPSRL"
               5223..5614
                /gene="Rv2562"
                /complement(5223..5614)
                /note="Rv2562, (MTCY9C4.06c), len: 129. Unknown"
                /codon_start=1
                /transl_table=1
                /product="hypothetical protein Rv2562"
                /protein_id="CAB01042.1"
                /db_xref="GI:1460078"
                /db_xref="SWISS-PROT:O50736"
                /translation="MAEOKVVRNVELACVDVILVHMLKNEVPSEVLFMTDVAOCL DESVKKLAPLTHDEPGICPSTHYIDLATSDMPRPAYVDHSHFGLINADVYEFEMHALP VILGFKKACAGFRSLRGATTEBPAMG"
                5757..6806
                 /gene="Rv2563"
                 /complement(5757..6806)
                 /note="Rv2563, (MTCY9C4.05c), len: 349. Unknown membrane protein, highly similar to A10214.28|MTV030_16

```

Mycobacterium tuberculosis seque (349 aa) fasta score:
 opt: 1772 z-score: 1874.0 E(1): 0.7628 identity in
 349 aa overlap

Query Match 60.0%; Score 15; DB 3; Length 15916;
 Best Local Similarity 100.0%; Pred. No. 3.2e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 gcaggtgcatgcgcca 20
 |||
 Db 735 GCAGGTGTCATGCCCA 749

RESULT 40
 AF065159
 LOCUS
 DEFINITION
 35209 bp DNA BCT 03-JAN-2000
 Bradyrhizobium japonicum putative arylsulfatase (arsa), putative
 soluble lytic transglycosylase precursor (slta),
 dihydrodipicolinate synthase (dapA), MscL (mscL), SmpB (smpB), and
 acyl carrier protein synthase Acps (acps) genes, complete cds;
 prokaryotic type I signal peptidase StpP (stpp) gene, stpP-sps
 allele, complete cds; RNase III (rnc), GTP-binding protein Era
 (era), hypothetical protein, recO-like protein, topoisomerase II
 (gyrA), putative ABC transporter, putative GMC-oxidoreductase,
 hypothetical protein, putative cytochrome c binding protein (cyc4),
 putative polytopic integral membrane protein, superoxide dismutase
 SodM-like protein, putative inner membrane protein, Atm-like
 protein (atm), adenylate cyclase-like protein, and putative inner
 membrane protein (spt) genes, complete cds, and unknown genes.
 AF065159.3 GI:6655022
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Bradyrhizobium japonicum.
 Bradyrhizobium japonicum
 Bacteria: Proteobacteria: alpha subdivision: Rhizobiaceae group:
 Bradyrhizobium group: Bradyrhizobium.
 1 (bases 12071 to 15486)
 Baill, A. and Muller, P.
 A second gene for type I signal peptidase in Bradyrhizobium
 japonicum, stpP, is located near genes involved in RNA processing
 and cell division
 Mol. Gen. Genet. 260 (4), 346-356 (1998)
 99066246
 2 (bases 1 to 12375)
 Mueller, P. and Stingel, D.
 Extended DNA sequencing in the upstream region of stpP in
 Bradyrhizobium japonicum
 Unpublished
 3 (bases 12071 to 15486)
 Baill, A. and Mueller, P.
 Direct Submission
 Submitted (13-MAY-1998) Biology, Philipps Universitaet Marburg,
 Karl-von-Frisch-Strasse, Marburg 35032, Germany
 4 (bases 1 to 35209)
 Mueller, P. and Stingel, D.
 Direct Submission
 Submitted (27-OCT-1999) Biology, Philipps Universitaet Marburg,
 Karl-von-Frisch-Strasse, Marburg 35032, Germany
 5 (bases 1 to 35209)
 Sequence update by submitter
 Mueller, P. and Stingel, D.
 Direct Submission
 Submitted (03-JAN-2000) Biology, Philipps Universitaet Marburg,
 Karl-von-Frisch-Strasse, Marburg 35032, Germany
 Sequence update by submitter
 On Jan 3, 2000 this sequence version replaced gi:6136295.
 location/Qualifiers
 1..35209
 /organism="Bradyrhizobium japonicum"
 /strain="USDA 110spc4"
 /db_xref="taxon:375"
 727..1647
 gene
 CDS
 gene
 CDS
 gene
 CDS

/gene="arsa"
 727..1647
 /gene="arsa"
 /note="similar to other arylsulfatases"
 /transl_table=1
 /codon_start=1
 /product="putative arylsulfatase"
 /protein_id="AF04318.1"
 /db_xref="GI:6136297"
 /translation="MGFTDFEGYLHLDMEDPAHANYPOSILATVGRNNVMSAT
 NVDDPTVQTRWKGITGOKIEDSGPLPKRMETVDDLELKITTEPITKAROKRPFVW
 LNPTRMHVTHLSEKYEENRNSENGWSVSGEPAQLDDIVGAVMKRLKDEGENDITV
 AFTTNGCAENETWPDGQTPFAGGKGTAMGGRFVPCMLRMPPGKAVAGVENGIMSGL
 DWPTTFVAAAGNPNINGEIQKQGLGDKTYKVIYLDGYPKPDGPDYRRRPVSPDLLEH
 GCHTERGAHRRIOIPLHRSERLVGRHGGC"
 complement(2043..3872)
 /gene="slta"
 complement(2043..3872)
 /note="similar to Rickettsia prowazekii strain Madrid E"
 /transl_table=1
 /product="putative soluble lytic transglycosylase
 precursor"
 /protein_id="AA04319.1"
 /db_xref="GI:6136298"
 /translation="MARPSDYRAFLISANSPSOTPLRRRLLEAMMDRRDSDVMSW
 FENSPYSAGKRFILAKAMLRQDRANAEPAGARGAQRSDGCHREORARPVRPAD
 AGREGADDTLSTAKTRPLAAKRLDAGVALAKARIASVKKAPNTALLLEVARTC
 TMPSGSEFLFSKIQLLRERKFAEAOLMSAPKDPRLVNLDEWMIERLLARKMID
 TEERSAVYLIAADALPSRDIVYTEDEFTGMLALRPLNDPATATQHAFRIGVSNP
 TTLRAGYWGGRAAEAAGROOEANNAVARAEOSTSYCOLARAKLGPILNNSQPR
 GRALNGRSCARPRLYLEDERMAVPLADMGENDPEALTGIGELTOYSPARGML
 LVGRSTPAAPRPDYAYPYNGIIPQPTGPEVRSRLRHPRAGERVOSTGLGAGG
 YGLKQVTPDARIVCKRHGATYDLGRKNSATINATYGSALGGLLEDYRGSYIMTPA
 RAYNAGRSVKKWMDYGDPPDPKVDADVWELLPESTRTNVAADHEPESGLPLRLR
 RHAVADGRPAVRPRRORGIEITALISIRREHRRMAPDMRRPA"
 4811..5701
 /gene="dapA"
 4811..5701
 /gene="dapA"
 /function="involved in the biosynthesis of the rhizopine
 3-O-methyl-scyllo-inosamine"
 /function="catalyzes the first step in the biosynthesis of
 diaminopimelate and lysine from aspartate semialdehyde"
 /note="similar to the Mosa protein of Sinorhizobium
 meliloti; belongs to the DHDPS family"
 /transl_table=1
 /codon_start=1
 /product="dihydrodipicolinate synthase"
 /protein_id="AA04320.1"
 /db_xref="GI:6136299"
 /translation="MAAKTKRGSFTALYPPFKNGSIDEAARFLYVMWQISEGNGLY
 PVGTTGESPTLSHDEHKVVEWCIIEANGRPVAVAGSSNTRKATLQHAERAGAD
 AYLVTTPYNNKPTQEGMQRKAINDIGPIIILYINIPRSVIDMVKRMLWELN
 IAGVADATASWVRVSOORAMGEDPNDISGEVATILIGYMQCGGICISVSNAPRLC
 SEPHRAMOKGDPORALKLHDKLMPVHNHNLFIENSPADIKYAMSLGLKDELTLRLPWP
 VPEPRVAVRSAMVHAGLIN"
 5729..6202
 /gene="mscL"
 5729..6202
 /note="similar to large conductance mechanosensitive
 channel proteins of other microorganisms; putative
 mechanosensitive channel protein large subunit; two
 transmembrane domains"
 /codon_start=1
 /transl_table=1
 /product="MscL"
 /protein_id="AA04321.1"
 /db_xref="GI:6136300"
 /translation="MSGVDEKGRRLKREFEFAMKGNVVDLAVGVIIIRARCAIVTSL
 VGDVIMPLIGAVTGLDFSNFTPLSKAVTATINLADAKKQCAVILAMGSEFLTLITINFIIT

```

gene
6236..6712
/ gene="smpB"
6236..6712
/ gene="smpB"
/ note="small protein homolog"
/ codon_start=1
/ transl_table=11
/ product="SmpB"
/ protein_id="AA04322.1"
/ db_xref="GI:6136301"
/ translation="MADKNERPIKVMENKARPNVAIEDTITAGIALTGTETKSTRN
GKSTIASYADSKNGETWLNATIPETLOGNRNHEPRKRLILHRKQILKIGVD
REGMTLPLEKLTNGRRAKLQLA VAGKRLHDKRETEKKRDWEPGRSACGARG"
6848..7297
/ gene="bcpB"
6848..7297
/ gene="bcpB"
/ note="bacterioferritin-comigrating protein"
/ codon_start=1
/ transl_table=11
/ product="BcpB"
/ protein_id="AA04323.1"
/ db_xref="GI:6136302"
/ translation="MLSALRGRTVVFAYPRTGEPKISLVDDMDMTPGARCGTPTQCA
FRLFAELKAAGAAYHGLSTOSNEYOTEMASLHLPPVLSPEKILALAKLPTWE
VAGLMLKRLALITIDAKITKFPYVFPEDRNAGDVLWLKANPVKD"
complement(7306..8157)
/ note="ORF283; contains two putative transmembrane
domains"
/ codon_start=1
/ transl_table=11
/ product="unknown"
/ protein_id="AA04324.1"
/ db_xref="GI:6136303"
/ translation="MIADRECAARPTSSPTSSSCNPSSAATRPNAPRRATVASAPAT
PRNSSARPRPMRGAMTTLSENAARGROPTLPDRDCLCRLVAFREGESRAV
VGIPTSPEDVDYARLLIVGLAPMGQANRGTPPGDYAGDLTLETLEPGAGV
QARPDDGLKIDCRIRANAVHCVPPKPKPAETINTCROFLANLAIATMLPRAIVAGR
IAHDTVLKPLMLKGSQAPFGHGAHVHQAQAFRLYDSYHCSRNTNTGVLTPDAPRSVFS
KYKADLD"
complement(8154..8603)
/ note="ORF149"
/ codon_start=1
/ transl_table=11
/ product="unknown"
/ protein_id="AA04325.1"
/ db_xref="GI:6136304"
/ translation="MSPTNKIALFIDGANLYATAKTLGPDIDYKRLLKEFQSGTL
LRATYTAILEDYSSIRPLIDMDYNGYTVATKATKEFIDASGRKRYKGMDEILA
VAMELAHIDDMVLFSSGDGDFRSIVEAVQGRGVTVSRSPASRR"
8929..9243
/ gene="rnpO"
8929..9243
/ gene="rnpO"
/ note="RNA-dependent nucleotide polymerase omega subunit
homolog"
/ codon_start=1
/ transl_table=11
/ product="RnpO"
/ protein_id="AA04326.1"
/ db_xref="GI:6136305"
/ translation="MARVTEPCIDKVDNRFDVLLAAHRAAMISSGSOLVYDRNDK
NPVSLREIADTTISPEDLRELGALAEVRRGRARAGYGAADRFRRRHSGCRYS
CR"
9484..11823
/ gene="relA"
9484..11823
/ gene="relA"
/ gene="relA"

```

Query Match 60.0%; Score 15; DB 1; Length 35209;
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Oy 7 cagtgatcgccag 21
|||||
Db 169 CAGTGATCGCCAG 183

RESULT 41
AC009603/c
LOCUS
DEFINITION
AC009603 35573 bp DNA INV 31-MAY-2000
Leishmania major chromosome 35 clone L3342 strain Friedlin,
complete sequence.
AC009603
AC009603.2 GI:6007874
HTG.
Leishmania major.
Leishmania major.
Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
Leishmania.
1 (bases 1 to 35573)
Myler,P.J., Sisk,E., Hixson,G., Kiser,P., Rickel,E., Hassebrock,M.,
Cawthra,J., Marsolini,F., Sunkin,S. and Stuart,K.D.
Direct Submission
Submitted (28-AUG-1999) Seattle Biomedical Research Institution, 4
Nickerson Street, Seattle, WA 98109-1651, USA
2 (bases 1 to 35573)
Myler,P.J., Sisk,E., Cawthra,J., Handley,F., Vogt,C., Robertson,L.,
McDonagh,P. and Stuart,K.
Direct Submission
Submitted (04-OCT-1999) Seattle Biomedical Research Institution, 4
Nickerson Street, Seattle, WA 98109-1651, USA
3 (bases 1 to 35573)
Myler,P.J. and Sunkin,S.
Direct Submission
Submitted (17-MAY-2000) Seattle Biomedical Research Institute, 4
Nickerson Street, Seattle, WA 98109-1651, USA
4 (bases 1 to 35573)
Myler,P.J.
Direct Submission
Submitted (31-MAY-2000) Seattle Biomedical Research Institute, 4
Nickerson Street, Seattle, WA 98109-1651, USA
On Oct 4, 1999 this sequence version replaced gi:580165.
Location/Qualifiers
1..35573
/organism="Leishmania major"
/strain="Friedlin"
/db_xref="taxon:3664"
/chromosome="35"
/clone="L3342"
<1..595
/ gene="STPK5"
/ note="L3342.2L; predicted using Glimmer, Testcode and
CodonUsage"
<1..595
/ gene="STPK5"
/ note="Blastp similarity to serine-threonine protein
kinases from several organisms; contains Eukaryotic
protein kinase domain"
/ codon_start=2
/ product="Serine Threonine Protein Kinase 5"
/ protein_id="AA070529.1"
/ db_xref="GI:7882847"
/ translation="IIALPPLPGSNELDQVRIHNVIGTPPEILERLKFGTHMY
DEPKQGTGKGLPHVSAFLDLMKRLITVDEORCTAKEARLHAAYFSKLRPADKKS
HRLKHSASTSRPTTVDDAASIGLSSSPKTTNGPPLNSTLTASRKLPVIDKSPPT
KSCITLVNOSTLHQVHGAATASAGDVLVARSPLKLI"
3863..4618
/ gene="L3342.5"
/ note="predicted using Glimmer, Testcode and CodonUsage;
CodonUsage not convincing at 5' end"
3863..4618
/ gene="L3342.5"
/ note="Tblastn similarity to several T. brucei GSS clones"
/ codon_start=1

```

/product="L3342.5"
/protein_id="AAF70530.1"
/db_xref="GI:7882848"
/translation="MHCPRGILGSGALPTIYADLTSCVSTIGRASPLGFSPT
PDKHPRIKQVGTGVAIOHVPGTTLKRTNKRGTFTSPVLAOLTHPPVRKATG
WHEIDTSDVDGDDDDMPSEVDECLAASVAGGELVINOELRALQOAFYALCD
VLMDTSGMKTSMYSSEINNGPDYELDRKTRRRHMLAIRHREYDKELLMSDAA
DDGASASVETASVNSVNRWQTSRCRCRLRT"
5035..6909
/gene="L3342.3"
/note="Predicted using Glimmer, Testcode and Codonusage;
ORF starts at 4615, Testcode & Codonusage predicts 3rd ATG
is used as start codon"
5035..6909
/gene="L3342.3"
/note="contains Glycican Pfam motif"
/codon_start=1
/product="L3342.3"
/protein_id="AAF70531.1"
/db_xref="GI:7882849"
/translation="MQICPEGSGDASSELLRESNDAGVYREYLDALCRVYPSL
FDALELVDAELCVLLRYATETQEDNAAFYVWGSEVMSRPERATTEGRIYLOH
FTLVLCARHLHHHDGPPPTLLSGDAENIPVLDQLYADHNHPEYVQKIGRANA
QGLRYAGLQSTASTMAELLRLSEANOPMAAGARGLDLRIHVAOKALGSKTL
EDVTRALPLLAOMGKTSEDOVKSTYERILTMSTTJAVLAOOPKPAQIALLEGIA
GCHYTPSSYKLEVMYLRMMRGAFSTLDAGRCILAMLOYGPKVQKVSOKATAMV
VLAADARSDDTAADVLAVALRISCTSSPASILGYSKSTPLMONTANWQSEHVAR
VLAASAAALETEANSEIGIAVSTTARERLGLSPHPPSPVAPHSFACVGEDA
CVTCAALDVSPPALTAAMAHADIDARVHSLCTPALAVSTMEALEMSIAISRMT
AVVYLDITLDEOMORPLEDDAETAVRCFLORAPGAVTRIASVLEHEVDRLDQTL
TGAIVVDDAARAPVCTEQRSRLRQTSSEGETRRLRYCALQHASREMLSTSAVY
7592..9346
/gene="L3342.4"
/note="Predicted using Glimmer, Testcode and Codonusage,
but Testcode and Codonusage not convincing, especially at
5' end"
7592..9346
/gene="L3342.4"
/codon_start=1
/product="L3342.4"
/protein_id="AAF70532.1"
/db_xref="GI:7882850"
/translation="MLRCETMTNAVESDVSSALRHSSNRSSTRATPLPGAAPR
GSSKRAVGMLEFSSAROOILKSKGNTVASTYGAADAVIDOTRQDTADAKYTFEL
DAPAMAHAMATTKKQSCNRPRLPGSTERTIVSRI SPRVALERHSRSTVTGMDYLT
RLQALNACDTSOSLPSPEDQVWQHRRPAMTAAALPCRRPRGRSGSGWQSVRST
TPSLGVTTLVKEWKAWEKATAVLTKGSTVECMESKSLTMAPFCMPSPRAGOLST
SCVARLOSKEFRPVASWMPDEPCASAGRDVAEEAOPWRRHLAEPLOSRLK
DAVSPVKPPPSLIVPVIGSARTRDMLERHSFEHQRILKQELGELDKLEMRHRS
ALLRHNSKEDAAVYPAHSPSTAPSTQGAERMRPPLAEVFGPSGSAARHOLY
ESPILYSLMGSITWSHEARAOVQEKSRVALLOSIVVTCASAVALTLAAEKIRSET
IRLVHWPENSGSHHAALLHREOKFLKLMGRYEAHTSGLLGAVDTVRPIDLIPSO
AEASFARAYAEVLPAAE"
10834..12516
/gene="PAB1"
/note="L3342.1: 511.1L; predicted using Glimmer, Testcode
and Codonusage; Glimmer predicts start codon at 10885"
10834..12516
/gene="PAB1"
/note="binds Poly(A) tail of mRNA; Blastp similarity to
many polyadenylate-binding proteins from several
organisms; contains RNA recognition motifs and bacterial
histidine kinase domain"
/codon_start=1
/product="PolyA Binding Protein 1"
/protein_id="AAF70533.1"
/db_xref="GI:7882851"
/translation="MAAAYGAAAPVAHOPQMDKPMQIATYVGDIDATINPOLVEL
FKPRTILNRYVCRDITITQSLGTYGVNFDNHSAEKATIESMNRKGDCKVRUMQO
RDPALRYSGNNGVFNKLEKVDKSLDHIITKGSILSCVWMDREKSGVGFVH
KDEITSADKAIKMGADHASEDKKALIVANFIRRNALALVANFTNTVYIKOVLPTV
NKDVLEKFFAFKGISSAACCKDKSGRVFAFCNFEKHDDAVKAVEAMHDDHIDITAP

GEKLYVQAPQSERERLIALROKYMHOALNNILYRNPDEPFTGADLLELEKEYGEVK
SCRVAVSESGVSGFGEVSEFNSADENALRMNGMNGKPLIYNIGRORPOATML
RLQFOORLOMMQMOHOPMPFUCSOGRRPGRGSGROOLCGRAGQMPMPROQPOAP
AOPQGRATPSAVGEVQATKRHSBGDVETPPPLPPIITPQLESMSSQEOBRALGDKLFL
KVEIAPLELAPKTTGMEKREKAEVALLNDKRLREERTEALCVLKAHQTA"
15582..17570
/gene="511.2"
/note="L3342.6: predicted using Glimmer, Testcode and
Codonusage; Glimmer predicts start codon at 15624"
15582..17570
/gene="511.2"
/codon_start=1
/product="511.2"
/protein_id="AAF70534.1"
/db_xref="GI:7882852"
/translation="MISNRYIAOHGSCHEVVPNGHEAVDAALGFRPPNAPMHTHT
VFRPSKNLTATFTTAASSSALGVNHTYQTLTPALANLVNHHGLRVDPSPRPPOD
AKSATSLPILPADRLSVGSDRLRGTEVEFRVAKPATNSSTSPNGAEGGDAA
AINGNHSLELVTSTAPASARQTESGOSTSALVPATLVPAKTAFASSDHCPSVY
SPSLRLVRKSTSTASTPPLQAPEVSAFTVATYLRLDVLMQORVLDKSOISGDA
RALPAPPAPEPPEAESQHOQTAVALIADTSTGSPANEDVMEDESDCYPLASTHAA
LVLSAERVRVGPVGAASPTFAAKYADVAVPDETRFVILAVSOPSAIMERRR
GPATRAPMBGLTSPNASHLEGALETGAKRTSRRITAEVPSAASVASTAQULTTAA
GOLQRSSPNAGGGEHPSLRAPAORGSAGSVSRHYHEDPSLPGIAEFLRAGNPSF
TSLSATHAAPPARAKASPTRRPRRSGPPAADSPAVLLGHSPILQRRGSSGTA
LSQADAPVPSLHLPYAHNEGATVADVSPRGSARARVDSPTPYAQPTLSLMSKGS
EVSADDFPMSRVASPTOTPLKDPVTASRDAASEPALL"
19700..21268
/gene="511.3"
/note="OrfAlpha; L3342.7: predicted using Glimmer,
Testcode and Codonusage"
19700..21268
/gene="511.3"
/note="Prosite predicts bipartite nuclear localization
signal"
/codon_start=1
/product="511.3"
/protein_id="AAF70535.1"
/db_xref="GI:7882853"
/translation="MLLSDBEAKITVAVLSLDDPTTYEVENLYONCERITAAVRAA
LADMSGDYVQDLCERCEAEKQVQAKRAGVASTAADAAFLDEFLAEBAQNEAEI
ESAISAAMHGRPRKKSQGEKSKKEGGERHKKRRLIDEPNPLPAPAGGSCEN
NAADKATSPSTRWVILSRDDLTSEVLPAALSTSDRVCAEVPNTLPAEFTEEVT
RDSASOCVAELTCVAVIAMRYGEPQLSMAECCKMQSOVAAASKAGTTLIEDGPPRS
AKAAGARSPTSSDQGBEGADSTSLYEMTSLBDLINOYQAOABEEBEDADTLT
LPRLISERFQRIHGFTEYVWLLCHGGTFAGGVARQCTTHKAFORTVVRKQGG
KQSSNKKDAGSYNSVGSQIRAVQDEVKRVADVRLDLMMPYITQAASTILYAAPGPQR
AVLTDSLLEPAVAVSGRGKGVSPITQKDPKRVSVPLTTRPTFEVRYRIVGCSRSL
LVYROEAOBRNRYN"
23224..23730
/gene="511.4"
/note="ORF2: L3342.8: predicted using Glimmer, Testcode
and Codonusage"
23224..23730
/gene="511.4"
/codon_start=1
/product="511.4"
/protein_id="AAF70536.1"
/db_xref="GI:7882854"
/translation="MSDYSDDDDVYVLOVCANRRCOGIDDEFEDETNOGYCGRCA
LYARTEGFRVLISNDQDGLVYLIDQDENSGRWTFYKEMNAFQEVYERSDDPTE
SSAALEFFFEEDYIDISAHGKGAVALLVLDENMYGGLVNNVADLVDESEAMEYQ

Query Match 60.0%; Score 15; DB 4; Length 35573;
Best local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 2 cgcagcagtgatc 16
Db 7912 CGCAGCAGTGATC 7898

RESULT 42


```

/protein_id="AA69577.1"
/db_xref="GI:7769949"
/translation="MSGSGATGSGSAGLSSSTINCKTSTASRSGNSTNNISITG
KKRSTTAAATTPACGSENGYHSSSTSPPTSDSEVASLTAAATAGSRKSPLT
PTPTIGATSCESAMTLHPAGTPARAKRPPATSLNLEVTALQALGAAPKREO
GDRVVESEFTIIVIAAGVKDETORTAETLISGADAAATASGNDGVAAPKSV
ATSDASALNTVSEFEEDTADLEITLEDIISNDPPEGCCIVNSPEKQISCTIM
YASEGYSASMSALVALDOKSYPPREDFLRCPRESERTALAPGSSISLPSTAT
DKTEDVLPYSYIFTHRYKYPPLGFIHHAUSLSTIEHRATNPALYADKEVUKMLM
ANADLDYLEDPTPLLSGLQCTAAPRSVPEPLRLHLIPEEYITETRDWVGF
LQMTVLPSCGAASVALLAKYATMSRVRNDIVCALICDSNFSACFALVLPQV
PAMASGSSARNRGLSVEASLVTSLSESGICSGMSGVNGASPSSTPAVDPVPAQ
KAELOVLAHLREFRELVTLSIAELGEMGPIYCKIYSGVSEPSISLVERAMPMT
AAAFVHMDGSGGALNGARNGGSSANSNAAGSGSTANCYSLAVEDEIAYL
DVTVLNROEQLMNEVRSPLADPKYKNITFPMRLQIVGCPITCAATRTSGTG
TNATRSSLTFPPPKILDSMSSTNPFALPHVLGHDEGSASERKLDSESTGRN
RFNFTITRYIAHATRGVMPVPLPSLGGVSGSTNPIIYOFNNANASONGARVH
ODNGSDTASATVGAHNTCPAAGGGSGAASVPLPGIITPALLIVLEMYVNTMA
NRELILKTEVHKSHIMHIEGVCDTLCENAGARCOIGLVLCGCVLEKVVYVYOM
TVTAEVDENPLSSVASSRREPLPOLYTTICQITVYERDTPGFIKAYNEIGRRDS
VYHSLISYDITIGKSVHPEDSTNNINRDVADYIFRHFTRLEPVEKCRREALS
EYTVRIGORATSPNDVSYSVLSTASSELTRKSSKURFVDEDAVSGVDGVVSVS
VSPALVWMAAEEKORRRHSGSMQMNATANIANGSPLOVRSASLGLADHGTGA
SPSSSPSRKSSPLTITIRPGSSNSSGSSGSSRTPAAGDLKVAAGHNSGVND
GPIAQSPPSTVPTTRKADYVLRPTAPKEKPPAPRNFASVLESPSTNAASGO
GKRSMPNATAKHAPPEKYLSTRTGSSVQSVGEHANGTALPESDNYVLPKIKTL
LSTGCKRS"
complement(2397..35102)
/gene="LA286.3"
/note="predicted using Glimmer, Testcode and Codonusage"
CDS
complement(32397..35102)
/gene="LA286.3"
/codon_start=1
/product="LA286.3"
/protein_id="AA69576.1"
/db_xref="GI:7769948"
/translation="MASASSAAVRAAARDNPSPPPPRTAORPRLVGLQCTP
SRSLKFSASMAHROFLAFDTAISRERFLFMELFEVROQLVDRVLPDPATYAP
BELLERAVITDPRHPAYGOAGLPAKSTIPSTIVISYSGTIEVPATSCSRPTTMGP
AYGDDVALAEFYNGATGRANDPRGVGTIOANLSAENRNNNGESTALVSRQINAG
BELIMSGAHRIASATPMVTVOGELIMRSLGCVLPFSFRETYESADSDTSAPAL
ATSLPORSASNAEDSAAEGSLTASVTDGASVSKDGSASAPADIDAGTLLS
ABDTKOPILARSRLVPCANDLMECTOCGMSICERSTADALRCACATPLSGARL
VLSLSPALTOAMLAEAGLDAGOCSTSTPOTOVSASVITSTSPSSRTMPM
NYPFLPMQVDAVPLVTLAKHSRETOHITVYDTCIDREKGRSVDHAGIGPT
VAAHPRGTPDAEQSGEACPSKSRPQPHSSADDSALAWVRDERPLVTEENAGR
GRGSAHSGSGSRNALMEROGDLDSELAITPGLIALPLLSVSVYSPAHALAS
AASPTKLPLEWTASAGERTSAGHSGVQGTSTALVSDTSDSVAGISTRPLS
YLSTIOEVRGLVLRRAASGSPPAAGTPYSPOMTIVRRLOCMTRIFLTKAPQD
DVSYVYGVLREGRDRCVNSCLEILPLRFLPRLRHRCGRGACIARSGSLCDA
EBAALOTPCDRIDRLSLVYTFNFMVCPCLVLEVTWEKNSVNAAGVEGSCADSVCYE
DEVDDALASCNVALVITMALCSPPFACAVAIKPIGAFELAKAO"
complement(36081..36680)
/gene="LA286.2"
/note="predicted using Glimmer, Testcode and Codonusage"
Query Match 60.0%; Score 15; DB 4; Length 41193;
Best Local Similarity 100.0%; Pred.No. 2.6e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 gcagcagtgatcatcg 17
|||||
Db 9371 GCAGCAGTGTCATCG 9385

```

```

SOURCE
ORGANISM
fruit fly.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
AUTHORS
Muzny, D.M., Adams, C., Bailey, M., Barbara, J., Blankenburg, K.,
Bodota, B., Bouck, J., Bowie, S., Brooks, A., Buhay, C., Bunn, C.,
Burkett, C., Burrows, J., Carter, M., Chacko, J., Chen, Z., Cox, C.,
David, R., Delgado, O., Deshazo, D., Ding, Y., Domah-Pashid, N.,
Dugan-Hoch, S., Durbin, K.J., Fernandez, C., Ferraguto, D.,
Forum-Tansey, J., Frantz, P., Ganesh, R., Gorrell, J.H., Gorrell, L.L.,
Guenera, W., Harris, K., Hernandez, J., Hodgson, A., Hogues, M.,
Holloway, C., Hoses, H., Jackson, L.E., Jackson, L., Jia, Y., Jones, M.,
Kelly, S., Kondrjewski, N., Kong, Y., Kovar, C., Leal, B., Li, Z.,
Lichtarge, O., Liu, J., Liu, W., Logan, O., Lozano, R.J., Lu, J.,
Lucier, R., Martin, R., Martinez, C., McLeod, M.P., Mei, G., Morgan, M.,
Morris, S., Nash, S., Nelson, A., Nguyen, R., Nguyen, N., Nguyen, S.,
Oswal, G., Parish, B., Paxton, S., Payton, B., Perez, L., Pu, L.L.,
Quiles, M., Reiter, D., Rives, M., Samuel, S., Say, J., Scherer, S.,
Shan, E., Shen, H., Simon, M., Sparks, A., Stamps, A., Sugang, R.,
Taber, P., Taylor, T., Vasquez, L., Vinson, R., Vo, Q., Wahbah, M.,
Watlington, S., Weinstock, G., Weinstock, I.R., Williamson, A.,
Worley, K., Wren, J., Wrensford, G., Yu, W., Zhou, X., Nelson, D. and
Gibbs, R.
Direct Submission
Unpublished
2 (bases 1 to 44545)
REFERENCE
AUTHORS
Worley, K.C.
TITLE
JOURNAL
Submitted (22-SEP-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Feb 18, 2000 this sequence version replaced gi:5917930.
COMMENT
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: DRHM
Center clone name: RPC198-10F16
----- Summary Statistics
Sequencing vector: M13; L08821
Chemistry: Dye-terminator Big Dye; 99% of reads
Assembly program: Phrap; version 0.980611
Consensus quality: 10931 bases at least Q40
Consensus quality: 20681 bases at least Q30
Consensus quality: 27126 bases at least Q20
Estimated insert size: 36762; sum-of-contigs estimation
Quality coverage: 0.4x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 32 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```

1 1244: contig of 1244 bp in length
* 1245 1264: gap of unknown length
* 1265 1280: contig of 1316 bp in length
* 1281 2600: gap of unknown length
* 2601 3779: contig of 1179 bp in length
* 3780 3799: gap of unknown length
* 3800 5212: contig of 1413 bp in length
* 5213 5232: gap of unknown length
* 5233 6780: contig of 1548 bp in length
* 6781 6800: gap of unknown length
* 6801 8288: contig of 1488 bp in length
* 8289 8309 9348: gap of unknown length
* 9348: contig of 1040 bp in length

```

```

* 9349 9368: gap of unknown length
* 9369 10616: contig of 1248 bp in length
* 10617 10636: gap of unknown length
* 10637 11240: contig of 604 bp in length
* 11241 11260: gap of unknown length
* 11261 12652: contig of 1392 bp in length
* 12653 12672: gap of unknown length
* 12673 14216: contig of 1544 bp in length
* 14217 14236: gap of unknown length
* 14237 15757: contig of 1521 bp in length
* 15758 15777: gap of unknown length
* 15778 17262: contig of 1485 bp in length
* 17263 17282: gap of unknown length
* 17283 18374: contig of 1092 bp in length
* 18375 18394: gap of unknown length
* 18395 19394: contig of 1000 bp in length
* 19395 19414: gap of unknown length
* 19415 20608: contig of 1194 bp in length
* 20609 20628: gap of unknown length
* 20629 21616: contig of 988 bp in length
* 21617 21636: gap of unknown length
* 21637 22529: contig of 893 bp in length
* 22530 22549: gap of unknown length
* 22550 23984: contig of 1435 bp in length
* 23985 24004: gap of unknown length
* 24005 24950: contig of 946 bp in length
* 24951 24970: gap of unknown length
* 24971 25858: contig of 888 bp in length
* 25859 25878: gap of unknown length
* 25879 27488: contig of 1610 bp in length
* 27489 27508: gap of unknown length
* 27509 29103: contig of 1595 bp in length
* 29104 29123: gap of unknown length
* 29124 30355: contig of 1232 bp in length
* 30356 30375: gap of unknown length
* 30376 31235: contig of 860 bp in length
* 31236 31255: gap of unknown length
* 31256 32692: contig of 1437 bp in length
* 32693 32712: gap of unknown length
* 32713 33672: contig of 960 bp in length
* 33673 33692: gap of unknown length
* 33693 35335: contig of 1643 bp in length
* 35336 35355: gap of unknown length
* 35356 36573: contig of 1218 bp in length
* 36574 36593: gap of unknown length
* 36594 38685: contig of 2092 bp in length
* 38686 38705: gap of unknown length
* 38706 41189: contig of 2484 bp in length
* 41190 41209: gap of unknown length
* 41210 44545: contig of 3336 bp in length.

```

```

FEATURES
  source
    1..44545
      /organism="Drosophila melanogaster"
      /db_xref="taxon:7227"
      /clone="RPC198-10F16"
BASE COUNT 11895 a 9996 c 9918 g 12034 t 702 others
ORIGIN

```

```

Query Match 60.0%; Score 15; DB 61; Length 44545;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 4 cagcaggtgcatgcc 18
  |||
Db 9470 CAGCAGTGCATGCC 9484

```

```

RESULT 44
AC014684/c AC014684 55307 bp DNA HTG 16-NOV-1999
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***; in ordered
pieces;

```

```

ACCESSION AC014684
VERSION AC014684.1 GI:6436651
KEYWORDS HTG; PHASE2.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
  Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
  Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
  Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
  1 (bases 1 to 55307)
REFERENCE Adams,M. and Venter,J.C.
  Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
  Rockville, MD, USA
  Direct Submission
  JOURNAL
  TITLE
  COMMENT
  This sequence was identified as CDM:10210542 by the submitter.
  For further information on this sequence e-mail to fly@celera.com.
  * NOTE: This is a 'working draft' sequence.
  * This sequence will be replaced
  * by the finished sequence as soon as it is available and
  * the accession number will be preserved.
  Location/Qualifiers
    source
      1..55307
        /organism="Drosophila melanogaster"
        /db_xref="taxon:7227"
BASE COUNT 16942 a 11043 c 11010 g 16312 t
ORIGIN

```

```

Query Match 60.0%; Score 15; DB 63; Length 55307;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 5 agcaggtgcatgcc 19
  |||
Db 6293 AGCAGTGCATGCC 6279

```

```

RESULT 45
AC014903 AC014903 73691 bp DNA HTG 16-NOV-1999
LOCUS Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***; in ordered
DEFINITION pieces.
ACCESSION AC014903
VERSION AC014903.1 GI:6436432
KEYWORDS HTG; PHASE2.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
  Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
  Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
  Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
  1 (bases 1 to 73691)
REFERENCE Adams,M. and Venter,J.C.
  Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
  Rockville, MD, USA
  Direct Submission
  JOURNAL
  TITLE
  COMMENT
  This sequence was identified as CDM:10210292 by the submitter.
  For further information on this sequence e-mail to fly@celera.com.
  * NOTE: This is a 'working draft' sequence.
  * This sequence will be replaced
  * by the finished sequence as soon as it is available and
  * the accession number will be preserved.
  Location/Qualifiers
    source
      1..73691
        /organism="Drosophila melanogaster"
        /db_xref="taxon:7227"
BASE COUNT 20452 a 16659 c 16681 g 19899 t
ORIGIN

```

```

FEATURES
  source
    1..73691
      /organism="Drosophila melanogaster"
      /db_xref="taxon:7227"
BASE COUNT 20452 a 16659 c 16681 g 19899 t
ORIGIN

```

```

Query Match 60.0%; Score 15; DB 63; Length 73691;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 2 cgcagcaggtgcatc 16

```

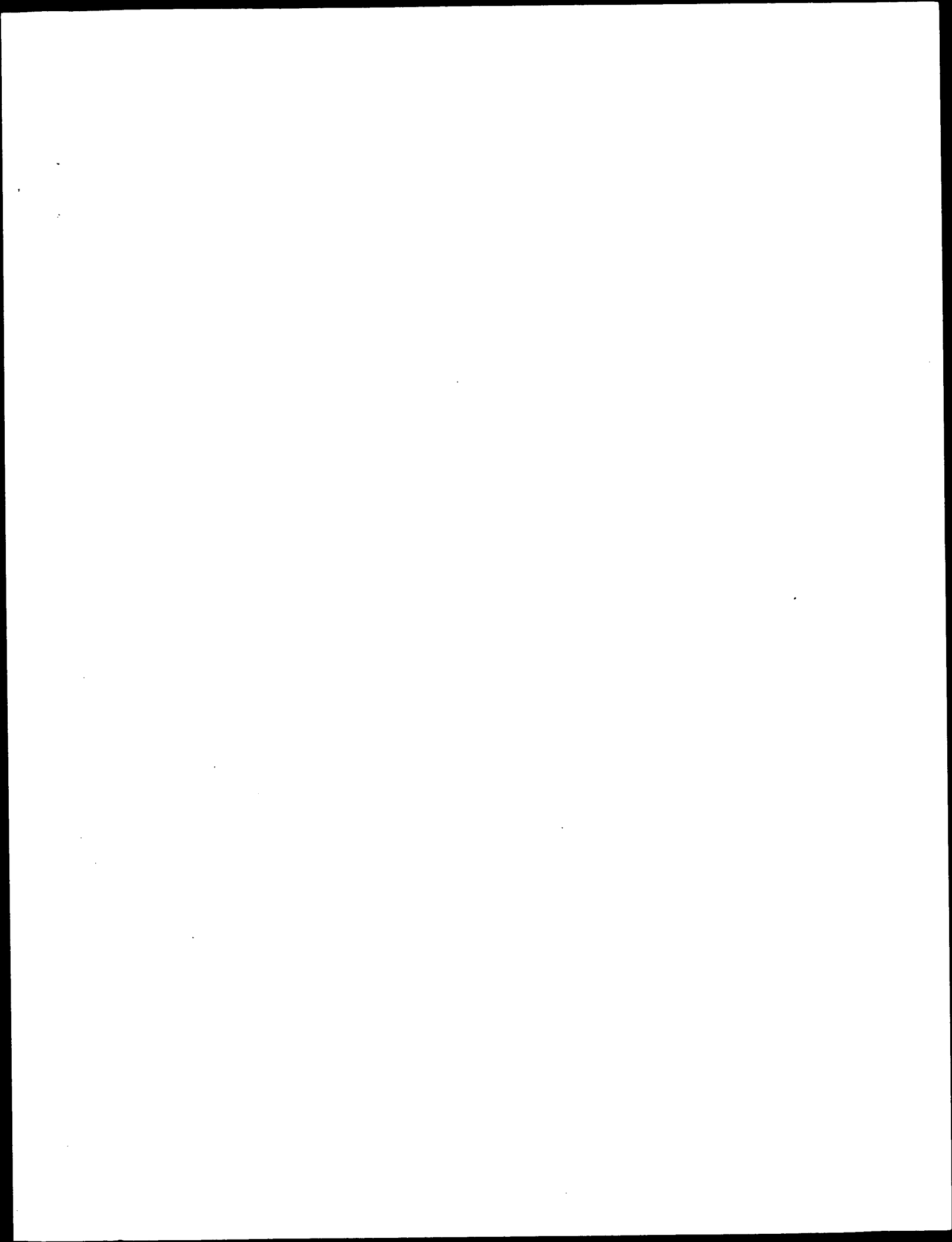

Wed Oct 10 07:45:52 2001

us-09-396-196f-5.oli.rge

Page 41

|||||||
Db 24505 CCGAGCGGTGCATC 24519

Search completed: October 9, 2001, 15:45:48
Job time: 14734 sec



XX Example 2: Column 37-40; 34pp; English.
PS

CC This sequence encodes the E. coli biotin synthetase (BioB). The gene can
CC be used in the transgenic plant of the invention. The transgenic plant,
CC plant cell or plant tissue is transformed with a chimeric gene encoding
CC diaminopelargonic acid (DAP) aminotransferase or biotin synthase and
CC produces more biotin than a non-transgenic plant, cell or tissue. The
CC plant is used as an improved dietary source of biotin (vitamin H) for
CC humans or animals.
XX

SQ Sequence 1041 BP; 262 A; 273 C; 305 G; 201 T; 0 other;

Query Match 100.0%; Score 25; DB 20; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.00017;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gccgcagcaggtgcatcgccagcatt 25
|||||
Db 79 gccgcagcaggtgcatcgccagcatt 103

RESULT 2

ID AAN91329 standard; DNA; 1084 BP.

AC AAN91329;

DT 15-FEB-1990 (first entry)

DE E. coli Bio B gene.

KW E. coli; Bio B gene; biotin.

OS Escherichia coli.

PH Key Location/Qualifiers
FT CDS 24..1064
/*tag=a

PN GB2216530-A.

PD 11-OCT-1989.

PF 17-MAR-1989; 89GB-0006210.

PR 22-MAR-1988; 88GB-0006804.

PR 17-MAR-1989; 89GB-0006210.

PA (UKAG-) UK MIN. AGRIC. FISH.

PI Pearson BM, McKee RA;

DR WPI; 1989-295085/41. P-PSDB P91392

XX Plasmid contg. gene(s) for expression of biotin synthetase enzymes
XX - derived from E. coli and capable of replication and expression in other
XX microorganisms, esp. yeast.
XX

PS Table 3; page 33-4; 52pp; English.

XX The gene can be used in a plasmid for expression of enzymes of the biotin
XX synthetic pathway. Pref. control sequences for expression in S. cerevisiae
XX are plasmids pMA91, pMA36c, pKV49 and pCK495, and plasmid pCK965 for
XX Lactobacillus. Insertion of bio B improves biotin yields in
XX microorganisms which export biotin, or enables growth in media contg.
XX little or no biotin of organisms unable to synthesise biotin for their
XX own use.
XX

SQ Sequence 1084 BP; 271 A; 286 C; 318 G; 209 T; 0 other;

Query Match 100.0%; Score 25; DB 10; Length 1084;
Best Local Similarity 100.0%; Pred. No. 0.00017;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gccgcagcaggtgcatcgccagcatt 25
|||||
Db 102 gccgcagcaggtgcatcgccagcatt 126

RESULT 3

ID AAN60496 standard; DNA; 1121 BP.

AC AAN60496;

DT 17-OCT-1991 (first entry)

DE Sequence encoding biotin synthesising enzyme.

KW Biotin synthetase enzyme; E. coli; desthiobiotin; ds.

PH Key Location/Qualifiers
FT CDS 42..1082
/*tag=a

PN JP61149091-A.

PD 07-JUL-1986.

PF 24-DEC-1984; 84JP-0272605.

PR 24-DEC-1984; 84JP-0272605.

PA (NIPS) NIPPON SODA KK.

DR WPI; 1986-216622/33.

DR P-PSDB; AAP60536.

XX Double stranded DNA encoding biotin synthesising enzyme -
XX comprises transformed mutant E. coli strain contg. cyclic doubled
XX stranded DNA encoding biotin synthetic biotin enzyme as a plasmid.
XX

PS Disclosure; Page 534; 23pp; Japanese.

XX The sequence may be expressed by a transformed E. coli host, cultured
XX in a medium containing desthiobiotin.
XX

SQ Sequence 1121 BP; 290 A; 301 C; 319 G; 211 T; 0 other;

Query Match 100.0%; Score 25; DB 7; Length 1121;
Best Local Similarity 100.0%; Pred. No. 0.00017;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gccgcagcaggtgcatcgccagcatt 25
|||||
Db 120 gccgcagcaggtgcatcgccagcatt 144

RESULT 4

ID AAO62386 standard; DNA; 5872 BP.

AC AAO62386;

DT 16-NOV-1994 (first entry)

DE Biotin-biosynthesis genes contg. plasmid pB030A-15/9.

KW Biotin; expression; enterobacteria; vitamin H; synthesis;

KW plasmid; pB030A-15/9; bioB; bioF; bioC; bioD; bioA;

KW promoter plac; biotin synthase; KAPA synthase;
KW 8-amino-7-oxononanoate synthase; pimeloyl-CoA; DTB synthase;

KW	dethiobiotin synthase: DAPA synthase:
KW	S-adenosyl-L-methionine: 8-amino-7-oxononanoate aminotransferase:
KM	seborrhoea; dermatitis: ds.
XX	
OS	Escherichia coli DSM498.
XX	
FH	Key
FT	promoter
FT	location/Qualifiers
FT	1..96
FT	/*tag= a
FT	/function= "promoter ptac"
FT	/evidence= EXPERIMENTAL
FT	23..28
FT	/*tag= b
FT	/standard_name= "promoter ptac"
FT	45..50
FT	/*tag= c
FT	/evidence= EXPERIMENTAL
FT	/standard_name= "promoter ptac"
FT	105..109
FT	/*tag= d
FT	/evidence= EXPERIMENTAL
FT	/standard_name= "biob RBS no. 9"
FT	117..1157
FT	/*tag= e
FT	/product= "biotin synthase"
FT	/evidence= EXPERIMENTAL
FT	/gene= "biob"
FT	/number= 1
FT	1141..1146
FT	/*tag= f
FT	/standard_name= "bioF RBS"
FT	1154..2311
FT	/*tag= g
FT	/EC_number= 2.3.1.47
FT	/product= "KAPA synthase"
FT	/evidence= EXPERIMENTAL
FT	/gene= "bioF"
FT	/number= 2
FT	/standard_name= "8-amino-7-oxononanoate synthase"
FT	2284..2288
FT	/*tag= h
FT	/standard_name= "bioc RBS"
FT	2295..3050
FT	/*tag= i
FT	/function= "involved in pimeloyl-CoA synthesis"
FT	/product= "protein"
FT	/gene= "bioc"
FT	/number= 3
FT	3030..3033
FT	/*tag= j
FT	/standard_name= "bioc RBS"
FT	3043..3753
FT	/*tag= k
FT	/EC_number= 6.3.3.3
FT	/product= "DTB synthase"
FT	/evidence= EXPERIMENTAL
FT	/gene= "biocD15"
FT	/number= 4
FT	/standard_name= "dethiobiotin synthase"
FT	3712..3750
FT	/*tag= l
FT	/note= "biocD15 substitution"
FT	3742..3746
FT	/*tag= m
FT	/standard_name= "bioc RBS"
FT	3750..5039
FT	/*tag= n
FT	/EC_number= 2.6.1.62
FT	/product= "DAPA synthase"
FT	/evidence= EXPERIMENTAL
FT	/gene= "bioc"
FT	/number= 5
FT	/standard_name= "S-adenosyl-L-methionine: 8-amino-

FT	RBS	5088..5093		7-Oxononanoate aminotransferase"
FT		/tag= O		
FT		/standard_name= "ORF1 RBS"		
FT	CDS	5098..5574		
FT		/tag= P		
FT		/function= "unknown, involved in biotin synthesis"		
FT		/product= "protein"		
FT		/evidence= EXPERIMENTAL		
FT		/gene= "ORF1"		
FT		/number= 6		
FT	terminator	5583..5644		
FT		//tag= q		
FT		/standard_name= "rho-independent transcriptional terminator"		
FT	stem_loop	5583..5605		
FT		//tag= r		
XX				
XX	W09408023-A.			
XX				
XX	14-APR-1994.			
XX				
XX	01-OCT-1993;	93WO-EP02688.		
XX				
PR	02-OCT-1992;	92CH-0003124.		
PR	15-JUL-1993;	93CH-0002134.		
XX				
PA	(LONZ) LONZA AG.			
XX				
PI	Birch O, Brass J, Fuhmann M, Shaw N;			
XX				
DR	WP1: 1994-135587/16.			
DR	P-PSDB; AAR51883, AAR51884, AAR51885, AAR51886, AAR51887, AAR63121.			
XX				
PT	Biotechnological biotin prodn. using enterobacterial biotin-gene			
PT	- providing vitamin H in high yield			
XX				
PS	Claim 1; Fig 6, Page 47-55 and 60-65; 92pp; German.			
XX				
CC	The sequence is derived from plasmid pB030A-15/9 contg. the			
CC	bioH, bioP, bioC, bioD and bioG genes responsible for biosynthesis			
CC	of biotin, arranged in a transcription unit. Microorganisms			
CC	contg. these DNA fragments or plasmids may be used in the prodn.			
CC	of biotin. Biotin (Vitamin H) may prevent seborrhoea, dermatitis,			
CC	loss of appetite and tiredness.			
XX				
SQ	Sequence 5872 BP; 1318 A; 1552 C; 1695 G; 1307 T; 0 other;			
<hr/>				
	Query Match	100.0%; Score 25; DB 15; Length 5872;		
	Best Local Similarity	100.0%; Pred. No. 0.00016;		
	Matches 25; Conservative	0; Mismatches 0; Indels 0; Gaps 0;		
OY	1 gcgcagcagtcgatcgccagcatt 25			
Db	195 gcgcagcagtcgatcgccagcatt 219			
<hr/>				
RESULT	5			
AAC51224	ID AAC51224 standard; DNA; 1498 BP.			
XX				
AC	AAC51224;			
XX				
DT	18-OCT-2000 (first entry)			
XX				
DE	Arabidopsis thaliana DNA fragment SEQ ID NO: 67727.			
XX				
KM	Hybridisation assay; genetic mapping; gene expression control;			
KW	Protein identification; signal transduction pathway;			
KW	metabolic pathway; promoter; termination sequence; ss.			
XX				
OS	Arabidopsis thaliana.			

XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 990S-0121825.
PR 05-MAR-1999; 990S-012180.
PR 09-MAR-1999; 990S-0125548.
PR 23-MAR-1999; 990S-0125788.
PR 25-MAR-1999; 990S-0126264.
PR 29-MAR-1999; 990S-0126785.
PR 01-APR-1999; 990S-0127462.
PR 06-APR-1999; 990S-0128234.
PR 08-APR-1999; 990S-0128714.
PR 16-APR-1999; 990S-0129845.
PR 19-APR-1999; 990S-0130077.
PR 21-APR-1999; 990S-0130449.
PR 23-APR-1999; 990S-0130510.
PR 23-APR-1999; 990S-0130891.
PR 28-APR-1999; 990S-0131449.
PR 30-APR-1999; 990S-0132048.
PR 30-APR-1999; 990S-0132407.
PR 04-MAY-1999; 990S-0132484.
PR 05-MAY-1999; 990S-0132485.
PR 06-MAY-1999; 990S-0132486.
PR 06-MAY-1999; 990S-0132487.
PR 07-MAY-1999; 990S-0132863.
PR 11-MAY-1999; 990S-0134256.
PR 14-MAY-1999; 990S-0134218.
PR 14-MAY-1999; 990S-0134219.
PR 14-MAY-1999; 990S-0134221.
PR 14-MAY-1999; 990S-0134370.
PR 18-MAY-1999; 990S-0134768.
PR 19-MAY-1999; 990S-0134941.
PR 20-MAY-1999; 990S-0135124.
PR 21-MAY-1999; 990S-0135353.
PR 24-MAY-1999; 990S-0135629.
PR 25-MAY-1999; 990S-0136021.
PR 27-MAY-1999; 990S-0136392.
PR 28-MAY-1999; 990S-0136782.
PR 01-JUN-1999; 990S-0137222.
PR 03-JUN-1999; 990S-0137528.
PR 04-JUN-1999; 990S-0137502.
PR 07-JUN-1999; 990S-0137724.
PR 08-JUN-1999; 990S-0138094.
PR 10-JUN-1999; 990S-0138540.
PR 10-JUN-1999; 990S-0138847.
PR 14-JUN-1999; 990S-0139119.
PR 16-JUN-1999; 990S-0139452.
PR 16-JUN-1999; 990S-0139453.
PR 17-JUN-1999; 990S-0139452.
PR 18-JUN-1999; 990S-0139454.
PR 18-JUN-1999; 990S-0139455.
PR 18-JUN-1999; 990S-0139456.
PR 18-JUN-1999; 990S-0139457.
PR 18-JUN-1999; 990S-0139458.
PR 18-JUN-1999; 990S-0139459.
PR 18-JUN-1999; 990S-0139460.
PR 18-JUN-1999; 990S-0139461.
PR 18-JUN-1999; 990S-0139462.
PR 18-JUN-1999; 990S-0139463.
PR 18-JUN-1999; 990S-0139750.
PR 18-JUN-1999; 990S-0139753.
PR 21-JUN-1999; 990S-0139817.
PR 22-JUN-1999; 990S-0139839.
PR 23-JUN-1999; 990S-0140353.
PR 23-JUN-1999; 990S-0140354.
PR 24-JUN-1999; 990S-0140695.
PR 28-JUN-1999; 990S-0140823.
PR 29-JUN-1999; 990S-0140991.
PR 30-JUN-1999; 990S-0141287.

PR 01-JUL-1999; 990S-0141842.
PR 01-JUL-1999; 990S-0142154.
PR 02-JUL-1999; 990S-0142055.
PR 06-JUL-1999; 990S-0142390.
PR 08-JUL-1999; 990S-0142803.
PR 09-JUL-1999; 990S-0142920.
PR 12-JUL-1999; 990S-0142977.
PR 13-JUL-1999; 990S-0143542.
PR 14-JUL-1999; 990S-0143624.
PR 15-JUL-1999; 990S-0144005.
PR 16-JUL-1999; 990S-0144085.
PR 16-JUL-1999; 990S-0144086.
PR 19-JUL-1999; 990S-0144325.
PR 19-JUL-1999; 990S-0144331.
PR 19-JUL-1999; 990S-0144332.
PR 19-JUL-1999; 990S-0144333.
PR 19-JUL-1999; 990S-0144334.
PR 19-JUL-1999; 990S-0144335.
PR 20-JUL-1999; 990S-0144352.
PR 20-JUL-1999; 990S-0144632.
PR 20-JUL-1999; 990S-0144814.
PR 21-JUL-1999; 990S-0144814.
PR 21-JUL-1999; 990S-0145086.
PR 22-JUL-1999; 990S-0145088.
PR 22-JUL-1999; 990S-0145085.
PR 22-JUL-1999; 990S-0145087.
PR 22-JUL-1999; 990S-0145089.
PR 22-JUL-1999; 990S-0145192.
PR 23-JUL-1999; 990S-0145145.
PR 23-JUL-1999; 990S-0145148.
PR 23-JUL-1999; 990S-0145224.
PR 26-JUL-1999; 990S-0145276.
PR 27-JUL-1999; 990S-0145913.
PR 27-JUL-1999; 990S-0145918.
PR 27-JUL-1999; 990S-0145919.
PR 28-JUL-1999; 990S-0145951.
PR 02-AUG-1999; 990S-0146386.
PR 02-AUG-1999; 990S-0146388.
PR 03-AUG-1999; 990S-0146389.
PR 04-AUG-1999; 990S-0147038.
PR 04-AUG-1999; 990S-0147204.
PR 05-AUG-1999; 990S-0147302.
PR 05-AUG-1999; 990S-0147192.
PR 06-AUG-1999; 990S-0147260.
PR 06-AUG-1999; 990S-0147303.
PR 06-AUG-1999; 990S-0147416.
PR 09-AUG-1999; 990S-0147493.
PR 10-AUG-1999; 990S-0147935.
PR 11-AUG-1999; 990S-0148171.
PR 12-AUG-1999; 990S-0148319.
PR 13-AUG-1999; 990S-0148341.
PR 13-AUG-1999; 990S-0148365.
PR 13-AUG-1999; 990S-0148684.
PR 16-AUG-1999; 990S-0149368.
PR 17-AUG-1999; 990S-0149175.
PR 18-AUG-1999; 990S-0149426.
PR 20-AUG-1999; 990S-0149722.
PR 20-AUG-1999; 990S-0148723.
PR 20-AUG-1999; 990S-0149929.
PR 23-AUG-1999; 990S-0149902.
PR 23-AUG-1999; 990S-0149930.
PR 25-AUG-1999; 990S-0150566.
PR 26-AUG-1999; 990S-0150884.
PR 27-AUG-1999; 990S-0151065.
PR 27-AUG-1999; 990S-0151066.
PR 27-AUG-1999; 990S-0151080.
PR 30-AUG-1999; 990S-0151303.
PR 31-AUG-1999; 990S-0151338.
PR 01-SEP-1999; 990S-0151930.
PR 07-SEP-1999; 990S-0152363.
PR 10-SEP-1999; 990S-0153070.
PR 13-SEP-1999; 990S-0153758.
PR 15-SEP-1999; 990S-0154018.

PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156658.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157865.
PR 06-OCT-1999; 99US-0158029.
PR 07-OCT-1999; 99US-0158332.
PR 08-OCT-1999; 99US-0158369.
PR 12-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 22-OCT-1999; 99US-0161044.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 25-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 64.0%; Score 16; DB 21; Length 1498;
Best Local Similarity 100.0%; Pred. No. 7.3;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 aggtgcacgcacgca 23
|||||||
Db 84 aggtgcacgcacgca 99

RESULT 6

AAC36258
ID AAC36258 standard; DNA: 1500 BP.

XX AAC36258;

XX 17-OCT-2000 (first entry)

DE Arabidopsis thaliana DNA fragment SEQ ID NO: 13129.

XX Hydridisation assay; genetic mapping; gene expression control;

KM protein identification; signal transduction pathway;

KM metabolic pathway; promoter; termination sequence; ss.

XX Arabidopsis thaliana.

XX EPI033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0128845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142194.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.

```

PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 23-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 09-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 10-AUG-1999; 99US-0147935.
PR 11-AUG-1999; 99US-0148171.
PR 12-AUG-1999; 99US-0148319.
PR 13-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0148684.
PR 17-AUG-1999; 99US-0149368.
PR 18-AUG-1999; 99US-0149175.
PR 20-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149723.
PR 23-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 25-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.

```

```

PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

Query Match Best Local Similarity 64.0%; Score 16; DB 21; Length 1500;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 8 aggtgcatgcgcacga 23
Db 84 aggtgcatgcgcacga 99

```

RESULT 7

AAFL1615/c
ID AAFL1615 standard; cDNA: 612 BP.

AAFL1615;

13-MAR-2001 (first entry)

Aspergillus niger EST SEQ ID NO:4138.

Multiple gene expression; filamentous fungal cell; EST;
expressed sequence tag; Fusarium venenatum; Aspergillus niger;
Aspergillus oryzae; Trichoderma reesei; identification; recombination;
culture condition; environmental stress; spore morphogenesis;
metabolic pathway engineering; catalytic pathway engineering; ss.

Aspergillus niger.

WO200056762-A2.

28-SEP-2000.

22-MAR-2000; 2000MO-US07781.

22-MAR-1999; 99US-0273623.

(NOVO) NOVO NORDISK BIOTECH INC.
(NOVO) NOVO NORDISK AS.

PI Berka RM, Rey MM, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
 XX WPI: 2000-594572/56.
 XX
 DR
 XX
 PT Monitoring differential expression of genes in filamentous fungal cells
 PT uses fluorescence-labeled nucleic acids isolated from the cells and a
 PT substrate of expressed sequence tags -
 PS
 XX
 PS Claim 87, Page 1813: 3161pp; English.
 XX
 CC The present invention describes a method for monitoring differential
 CC expression of genes in a first filamentous fungal (FF) cell relative to
 CC expression of the same genes in one or more second filamentous fungal
 CC cells. The method uses fluorescence-labeled nucleic acids isolated from
 CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
 CC are used in the methods for monitoring differential expression of genes
 CC in a first filamentous fungal (FF) cell relative to expression of the
 CC same genes in one or more second filamentous fungal cells. Monitoring
 CC the global expression of genes from FF cells allows the production
 CC potential of the microorganisms to be improved. New genes may be
 CC discovered, possible functions of unknown open reading frames can be
 CC identified and gene copy number variation and stability can be
 CC monitored. The expression of genes can be used to study how FF cells
 CC adapt to changes in culture conditions, environmental stress, spore
 CC morphogenesis, recombination, metabolic or catabolic pathway
 CC engineering. Using ESTs provides several advantages over genomic or
 CC random cDNA clones including elimination of redundancy as one spot on an
 CC array equals one gene or open reading frame, and organisation of the
 CC microarrays based on function of the gene products to facilitate
 CC analysis of the results. AAF07478 to AAF11247 represents ESTs from
 CC *Pisarium venetum*; AAF11248 to AAF1853 represents ESTs from *Aspergillus*
 CC *niger*; AAF11854 to AAF14878 represents ESTs from *Aspergillus oryzae*; and
 CC AAF14879 to AAF15337 represents ESTs from *Trichoderma reesei*, which are
 CC all specifically claimed in the present invention.
 XX
 SQ Sequence 612 BP; 147 A; 162 C; 144 G; 156 T; 3 other;

Query Match 60.0%; Score 15; DB 21; Length 612;
 Best Local Similarity 100.0%; Pred. No. 24;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 gcagcaggtcatcg 17
 |||||
 DB 118 GCACGACGTGCATCG 104

RESULT 8
 AAC93313
 ID AAC93313 standard; cDNA; 2597 BP.
 XX
 AC AAC93313;
 XX
 DT 16-FEB-2001 (first entry)
 XX
 DE Human secreted protein cDNA sequence #4.
 XX
 KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
 KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; anticancer;
 KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;
 KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
 KW neurological disease; infection; human; secreted protein; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200058495-A1.
 XX
 PD 05-OCT-2000.
 XX
 PF 23-MAR-2000; 2000WO-US07661.
 XX
 PR 26-MAR-1999; 99US-0126504.
 PR 07-JAN-2000; 2000US-0174847.

XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Ruben SM, Komatsoulis G;
 XX
 DR WPI: 2000-611720/58.
 DR P-PSDB: AAB51383.
 XX
 PT New nucleic acid molecules encoding 45 human secreted proteins for
 PT diagnosing, preventing, treating or ameliorating medical conditions and
 PT used as food additives or preservatives -
 PS
 XX
 PS Claim 1, Pages 327-328; 410pp; English.
 XX
 CC The invention relate to the isolation of genes AAC93310-C93354 encoding
 CC 45 human secreted proteins AAB51380-B51423. The genes can be used to
 CC generate fusion proteins by linking to the gene for the human
 CC immunoglobulin G Fc portion (AAC93301) for increasing the stability of
 CC the fusion protein as compared to the human protein only. The genes and
 CC proteins are useful for preventing, ameliorating or treating medical
 CC conditions, e.g. by protein or gene therapy. The genes are isolated
 CC from a range of human tissues disclosed in the specification. The
 CC nucleic acids, proteins, antibodies and (ant)agonists are useful in
 CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
 CC and ovarian cancer, and other cancers of the adrenal gland, bone, bone
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
 CC (b) immune disorders e.g. Addison's disease, diabetes mellitus, Crohn's
 CC haemolytic anemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d)
 CC wound healing; (e) neurological diseases e.g. cerebral anoxia and
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
 CC and parasitic infections.
 XX
 SQ Sequence 2597 BP; 721 A; 604 C; 534 G; 737 T; 1 other;

Query Match 60.0%; Score 15; DB 21; Length 2597;
 Best Local Similarity 100.0%; Pred. No. 24;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 gcagcaggtcatcgcca 20
 |||||
 DB 2256 gcagcaggtcatcgcca 2270

RESULT 9
 AAA01491/C
 ID AAA01491 standard; cDNA; 300 BP.
 XX
 AC AAA01491;
 XX
 DT 19-MAY-2000 (first entry)
 XX
 DE Human colon cancer cell line polynucleotide sequence SEQ ID NO:1482.
 XX
 KW Human; colon cancer; tumour; diagnosis; gene expression product;
 KW probe; detection; cancerous state; metastasis; identification;
 KW breast cancer; oestrogen receptor; positive breast cancer; therapy;
 KW oestrogen receptor-negative breast cancer; lung cancer; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO958675-A2.
 XX
 PD 18-NOV-1999.
 XX
 PF 13-MAY-1999; 99WO-US10602.
 XX
 PR 14-MAY-1998; 98US-0085426.
 PR 15-MAY-1998; 98US-0085337.
 PR 15-MAY-1998; 98US-0085696.
 PR 21-OCT-1998; 98US-0105234.

PR 27-OCT-1998; 98US-0105877.
 XX (CHIR) CHIRON CORP.
 PA (HYSE-) HYSEQ INC.
 XX
 PI Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
 PI Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;
 PI Lanson G, Dzmanac R, Cirvenjakov R, Dickson M, Dzmanac S, Labat I;
 PI Leshkowitz D, Kita D, Garcia V, Jones LM, Stache-Crain B;
 XX WPI; 2000-126369/11.
 DR
 XX Polynucleotide library used to determine cancerous states of mammalian
 PT cells -
 PI
 XX Claim 1; Page 536; 1097Pp; English.
 XX
 CC AAA00010 to AAA02716 represent polynucleotides isolated from cDNA
 CC libraries constructed from human colon cancer cell lines. The present
 CC invention also describes a method of detecting differentially expressed
 CC genes correlated with a cancerous state of a mammalian cell, comprising
 CC detecting at least one differentially expressed gene product in a test
 CC sample derived from a cell suspected of being cancerous, where detection
 CC of the differentially expressed gene product is correlated with a
 CC cancerous state of the cell from which the test sample was derived.
 CC The polynucleotides sequences can be used in a method for detecting
 CC differentially expressed genes correlated with a cancerous state of a
 CC mammalian cell. The polynucleotides can also be used as probes for
 CC detecting and mapping related genes. They can be used in diagnosis and
 CC prognosis of diseases and disorders (e.g. identification of
 CC pre-metastatic or metastatic cancerous states, stages of cancer, or
 CC responsiveness of cancer to therapy). This is particularly for breast
 CC cancer, oestrogen receptor-positive breast cancer, oestrogen receptor-
 CC negative breast cancer, lung cancer, and colon cancer.
 XX
 SO Sequence 300 BP; 56 A; 89 C; 94 G; 61 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 300;
 Best Local Similarity 100.0%; Pred. No. 81;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 gcagcaggtgcatc 16
 ||||||||||||
 DB 108 GCAGCAGGTGCATC 95

RESULT 10
 AAF71326
 ID AAF71326 standard; DNA; 667 BP.
 XX
 AC AAF71326;
 XX
 DT 30-APR-2001 (first entry)
 XX
 DE Corynebacterium glutamicum HA protein nucleotide sequence SEQ ID NO:377.
 XX
 KW Corynebacterium glutamicum; homeostasis; adaptation; HA protein;
 KW fine chemical production; organic acid; proteinogenic amino acid;
 KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;
 KW nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;
 KW carbohydrate; aromatic compound; vitamin; cofactor; polypeptide; enzyme;
 KW diagnosis; Corynebacterium diphtheriae; genetic engineering;
 KW Brevibacterium; environmental condition; ds.
 XX
 XX Corynebacterium glutamicum.
 OS
 XX
 PN WO200100842-A2.
 XX
 PD 04-JAN-2001.
 XX
 PF 23-JUN-2000; 2000WO-IB00911.
 XX

PR 25-JUN-1999; 99US-0141031.
 PR 08-JUL-1999; 99DE-1031636.
 PR 09-JUL-1999; 99DE-1032125.
 PR 09-JUL-1999; 99DE-1032126.
 PR 09-JUL-1999; 99DE-1032127.
 PR 09-JUL-1999; 99DE-1032128.
 PR 09-JUL-1999; 99DE-1032129.
 PR 09-JUL-1999; 99DE-1032226.
 PR 14-JUL-1999; 99DE-1032920.
 PR 14-JUL-1999; 99DE-1032922.
 PR 14-JUL-1999; 99DE-1032924.
 PR 14-JUL-1999; 99DE-1032928.
 PR 14-JUL-1999; 99DE-1032930.
 PR 14-JUL-1999; 99DE-1032933.
 PR 14-JUL-1999; 99DE-1032935.
 PR 14-JUL-1999; 99DE-1032973.
 PR 14-JUL-1999; 99DE-1033002.
 PR 14-JUL-1999; 99DE-1033003.
 PR 14-JUL-1999; 99DE-1033005.
 PR 14-JUL-1999; 99DE-1033006.
 PR 31-AUG-1999; 99DE-1041378.
 PR 31-AUG-1999; 99DE-1041379.
 PR 31-AUG-1999; 99DE-1041390.
 PR 03-SEP-1999; 99DE-1041391.
 PR 03-SEP-1999; 99DE-1042088.
 XX
 PA (BADI) BASF AG.
 XX
 XX Pompejus M, Kroege B, Schroeder H, Zelder O, Haberhauer G;
 PI WPI; 2001-061974/07.
 DR P-PSDB; AAB79211.
 XX
 PT New isolated Corynebacterium glutamicum nucleic acid for production or
 PT modulation of production of fine chemicals such as amino acids,
 PT nucleosides, nucleotides, lipids, fatty acids, carbohydrates, vitamins
 PT or enzymes -
 XX
 PS Claim 3; Page 630-631; 712Pp; English.
 XX
 CC AAF71138 to AAF71357 encode the Corynebacterium glutamicum homeostasis
 CC and adaptation (HA) proteins given in AAB79023 to AAB79242. The
 CC C. glutamicum HA genes (I) can be used in vectors for expression in host
 CC cells and production of fine chemicals, such as, an organic acid,
 CC proteinogenic or nonproteinogenic amino acid (preferred), purine or
 CC pyrimidine base, nucleoside, nucleotide, lipid, saturated or unsaturated
 CC fatty acid, diol, carbohydrate, aromatic compound, vitamin, cofactor,
 CC polypeptide or enzyme. The amino acids produced can be lysine, glutamine,
 CC glutamate, alanine, aspartate, glycine, serine, threonine, methionine,
 CC cysteine, valine, leucine, isoleucine, arginine, proline, histidine,
 CC tyrosine, phenylalanine, or tryptophan. The fine chemical production can
 CC be modulated. The presence of (I) or HA proteins encoded by then are
 CC used for diagnosing the presence or activity of Corynebacterium
 CC diphtheriae. (I) can be used to map the C. glutamicum genome or can be
 CC used as markers for genetically engineered Corynebacterium or
 CC Brevibacterium. The HA proteins encoded by the (I) are used to maintain
 CC homeostasis in C. glutamicum or help the microorganism to adapt to
 CC different environmental conditions.
 XX
 SO Sequence 667 BP; 146 A; 210 C; 172 G; 139 T; 0 other;

Query Match 56.0%; Score 14; DB 22; Length 667;
 Best Local Similarity 100.0%; Pred. No. 80;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gcgcagcaggtgca 14
 ||||||||||||
 DB 168 gcgcagcaggtgca 181

RESULT 11
 AAV20504/c

```

ID  AAV20504 standard; DNA; 719 BP.
XX
XX  AAV20504;
XX
XX  30-JUN-1998 (first entry)
XX
XX  Conus geographus conantokin DNA.
DE
XX  Conantokin; predatory cone snail; treatment: neurologic disorder;
XX  psychiatric disorder; anticonvulsant; neuroprotective; analgesic;
XX  HIV infection; ophthalmic indication; memory; learning defect;
XX  cognitive defect; ss.
XX
XX  Conus geographus.
OS
XX
XX  Key      Location/Qualifiers
XX  CDS      110..412
XX           /*tag= a
XX           /product= conantokin
XX
XX  MO9803541-A1.
XX
XX  29-JAN-1998.
XX
XX  21-JUL-1997; 97WO-US12618.
XX
XX  22-JUL-1996; 96US-0684742.
XX
XX  (COGN-) COGNETIX INC.
XX  (UTAH ) UNIV UTAH RES FOUND.
XX
XX  Abogadie FC, Colledge C, Cruz LJ, Hillyard DR, Jimenez E;
XX  Layer RT, McCabe RT, Olivera BM, Rivier JE, Shen GS;
XX  Walker C, Zhou L;
XX
XX  WPI; 1998-120694/11.
XX  P-PSDB; AAM48210.
XX
XX  New conantokin peptide(s) - useful for e.g. treating neurologic or
XX  psychiatric disorders, or the management of pain
XX
XX  Claim 19; Pages 79-80; 122pp; English.
XX
XX  The present sequence encodes Conus geographus conantokin, peptide
XX  derivatives of which can be used to treat neurologic and
XX  psychiatric disorders, e.g. as an anticonvulsant, neuroprotective
XX  or analgesic agent. Neurologic and psychiatric disorders include
XX  epilepsy, convulsions, neurotoxic injury (associated with
XX  conditions of hypoxia, anoxia or ischaemia, which typically follow
XX  stroke, cerebrovascular accident, brain or spinal cord trauma,
XX  myocardial infarct, physical trauma, drowning, suffocation,
XX  perinatal asphyxia or hypoglycaemic events), neurodegeneration
XX  (associated with Alzheimer's disease, senile dementia, Amyotrophic
XX  lateral Sclerosis, Multiple Sclerosis, Parkinson's disease,
XX  Huntington's disease, Down's Syndrome, Korsakoff's disease,
XX  schizophrenia, AIDS dementia, multi-infarct dementia, Binswanger
XX  dementia and neuronal damage associated with uncontrolled
XX  seizures), chemical toxicity (such as addiction, and morphine,
XX  opiate, opioid and barbiturate tolerance), pain (acute, chronic,
XX  migraine), anxiety, major depression, manic-depressive illness,
XX  obsessive-compulsive disorder, schizophrenia and mood disorders
XX  (such as bipolar disorder, unipolar depression, dysthymia and
XX  seasonal affective disorder) and dystonia (movement disorder),
XX  sleep disorder, muscle relaxation and urinary incontinence. The
XX  peptide can also be used to treat HIV infection, ophthalmic
XX  indication and memory, learning or cognitive defects.
XX
XX  Sequence 719 BP; 217 A; 171 C; 149 G; 182 T; 0 other;

```

```

OY  7 caggtgcatcgcca 20
DB  118 CAGGTGTCATCGCCA 105

RESULT 12
AAV17130/C
ID  AAV17130 standard; DNA; 719 BP.
XX
XX  AAV17130;
XX
XX  30-JUN-1998 (first entry)
XX
XX  Conus geographus conantokin DNA.
DE
XX
XX  Conantokin; predatory cone snail; treatment: neurologic disorder;
XX  psychiatric disorder; anticonvulsant; neuroprotective; analgesic;
XX  HIV infection; ophthalmic indication; memory; learning defect;
XX  cognitive defect; ss.
XX
XX  Conus geographus.
OS
XX
XX  Key      Location/Qualifiers
XX  CDS      110..412
XX           /*tag= a
XX           /product= conantokin
XX
XX  MO9803189-A1.
XX
XX  29-JAN-1998.
XX
XX  21-JUL-1997; 97WO-US12652.
XX
XX  06-DEC-1996; 96US-0762377.
XX  22-JUL-1996; 96US-0684750.
XX
XX  (COGN-) COGNETIX INC.
XX
XX  Layer RT, McCabe RT, Zhou L;
XX
XX  WPI; 1998-120469/11.
XX  P-PSDB; AAM49989.
XX
XX  Use of conantokin peptide(s) - for treating disorders involving
XX  excessive excitation of nerve cells by excitatory amino acids or
XX  agonists of the N-methyl-D-aspartate receptor
XX
XX  Example 4; Pages 79-80; 122pp; English.
XX
XX  The present sequence encodes Conus geographus conantokin, peptide
XX  derivatives of which can be used to treat neurologic and
XX  psychiatric disorders, e.g. as an anticonvulsant, neuroprotective
XX  or analgesic agent. Neurologic and psychiatric disorders include
XX  epilepsy, convulsions, neurotoxic injury (associated with
XX  conditions of hypoxia, anoxia or ischaemia, which typically follow
XX  stroke, cerebrovascular accident, brain or spinal cord trauma,
XX  myocardial infarct, physical trauma, drowning, suffocation,
XX  perinatal asphyxia or hypoglycaemic events), neurodegeneration
XX  (associated with Alzheimer's disease, senile dementia, Amyotrophic
XX  lateral Sclerosis, Multiple Sclerosis, Parkinson's disease,
XX  Huntington's disease, Down's Syndrome, Korsakoff's disease,
XX  schizophrenia, AIDS dementia, multi-infarct dementia, Binswanger
XX  dementia and neuronal damage associated with uncontrolled
XX  seizures), chemical toxicity (such as addiction, and morphine,
XX  opiate, opioid and barbiturate tolerance), pain (acute, chronic,
XX  migraine), anxiety, major depression, manic-depressive illness,
XX  obsessive-compulsive disorder, schizophrenia and mood disorders
XX  (such as bipolar disorder, unipolar depression, dysthymia and
XX  seasonal affective disorder) and dystonia (movement disorder),
XX  sleep disorder, muscle relaxation and urinary incontinence. The
XX  peptide can also be used to treat HIV infection, ophthalmic
XX  indication and memory, learning or cognitive defects.

```

```

Query Match      56.0%; Score 14; DB 19; Length 719;
Best Local Similarity 100.0%; Pred. No. 80;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

XX Sequence 719 BP; 217 A; 171 C; 149 G; 182 T; 0 other;
 SQ

Query Match 56.0%; Score 14; DB 19; Length 719;
 Best Local Similarity 100.0%; Pred. NO. 80;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 cagtgcatcgcca 20
 |||||
 DB 118 CAGTGCAATCGCCA 105

RESULT 13

AAC98831
 ID AAC98831 standard; cDNA; 1190 BP.

AC AAC98831;

DT 09-MAR-2001 (first entry)

DE Human pancreatic cancer antigen nucleotide sequence SEQ ID NO:59.

KW Human; pancreas; pancreatic cancer; pancreatic cancer antigen;
 KW detection; diagnosis; identification; cytostatic; neuroprotective;
 KW neotropic; immunomodulatory; relaxant; contraceptive; gynaecological;
 KW antinflammatory; cardiant; gene therapy; chromosome mapping;
 KW linkage analysis; tissue identification; tissue typing; forensic;
 KW neural; immune system; muscular; reproductive; gastrointestinal;
 KW pulmonary; cardiovascular; renal; proliferative; ss.

OS Homo sapiens.

PN WO200055320-A1.

PD 21-SEP-2000.

PF 08-MAR-2000; 2000WO-US05989.

PR 12-MAR-1999; 99US-0124270.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Ruben SM;

WP: 2000-579444/54.

P-PSDB; AAB54066.

XX New nucleic acid that is a pancreatic cancer antigen for preventing,
 PT treating, or ameliorating a medical condition, particular pancreatic
 PT cancer, or for use in assays for diagnosing a pathological condition -
 XX
 PS Claim 1; Page 537; 1379pp; English.

XX AAC98773 to AAC99231 encode the human pancreatic cancer associated
 CC proteins, called pancreatic cancer antigens, given in AAB54008 to
 CC AAB54466. The human pancreatic cancer antigens have cytostatic,
 CC neuroprotective, neotropic, immunomodulatory, relaxant, contraceptive,
 CC gynaecological, cardiant and antinflammatory activities, and can be used
 CC in gene therapy. The polynucleotide and proteins can be used for
 CC preventing, treating, or ameliorating a medical condition or in assays
 CC for diagnosing a pathological condition or a susceptibility to one in a
 CC subject. Binding partners to the proteins and the activity of the
 CC proteins can be identified. The pancreatic cancer antigens can be used to
 CC detect, treat or prevent pancreatic disorders, especially cancer.
 CC Agonists and antagonists to the antigens can be screened for. The
 CC pancreatic cancer antigen polynucleotides can be used to design nucleic
 CC acid hybridisation probes that can be used in chromosome mapping, linkage
 CC analysis, tissue identification and/or typing and a variety of forensic
 CC and diagnostic methods. The proteins can be used to generate antibodies
 CC which are used to purify, detect and target the polypeptides, including
 CC both in vivo and in vitro diagnostic and therapeutic methods. The
 CC proteins can be used to treat or prevent neural, immune system, muscular,

CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
 CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
 CC sequences used in the exemplification of the present invention.
 XX

SQ Sequence 1190 BP; 269 A; 323 C; 329 G; 267 T; 2 other;

Query Match 56.0%; Score 14; DB 21; Length 1190;
 Best Local Similarity 100.0%; Pred. NO. 79;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 ggtgcatcgccagc 22
 |||||
 DB 1141 ggtgcatcgccagc 1154

RESULT 14

AAZ56381/C
 ID AAZ56381 standard; DNA; 1344 BP.

AC AAZ56381;

DT 17-MAR-2000 (first entry)

DE Escherichia coli flagellin protein nucleotide sequence SEQ ID NO:51.

KW Flagellin; flc; antigen; detection; ds.

OS Escherichia coli.

PN WO961458-A1.

PD 02-DEC-1999.

PF 21-MAY-1999; 99WO-AU00385.

PR 21-MAY-1998; 98AU-0003634.

PA (UNSY) UNIV SYDNEY.

PI Reeves PR, Wang L;

WP: 2000-072598/06.

PT Novel nucleic acid molecule useful for the detection of flagellated
 PT bacterial strains in food, faeces, etc. -
 XX

PS Claim 3; Page 225; 245pp; English.

XX AAZ56331 to AAZ56398 represent nucleic acid molecules (1) encoding all
 CC or part of an Escherichia coli flagellin protein except a protein
 CC expressed by E. coli H1, H7, H12 or H48 type strains. The present
 CC invention also describes a method of detecting the presence of E. coli
 CC of a particular H serotype in a sample, comprising specifically
 CC hybridising a nucleic acid, preferably at least a pair, derived from a
 CC flagellating gene, specific for a particular flagellin gene associated
 CC with the H serotype, to any E. coli in the sample which contain the gene,
 CC and detecting any hybridised molecules, identifying the presence of that
 CC serotype in the sample. (1) are useful for: (1) detecting the presence
 CC of E. coli of H serotype in a sample by hybridising at least one or a
 CC pair of (1) to any E. coli in the sample and detecting the hybridised
 CC nucleic acid molecules; and (2) for detecting the presence of both O
 CC and H-serotypes of E. coli by hybridising at least one or a pair of (1)
 CC to any E. coli present in the sample and detecting the hybridised
 CC nucleic acid molecules. (1) is particularly useful for detecting the
 CC combination of O and H antigen. Hybridised (1) when using at least one
 CC (1) is detected by southern blot analysis and, when using a pair of (1),
 CC is detected by polymerase chain reaction (PCR). AAZ56399 to AAZ56420
 CC represent primers used in the exemplification of the present invention.
 XX

SQ Sequence 1344 BP; 362 A; 337 C; 322 G; 323 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 1344;
 Best Local Similarity 100.0%; Pred. No. 79;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 gcatcgccagcatt 25
 |||||
 Db 590 GCATCGCCAGCATT 577

RESULT 15

AAA69762/c
 ID AAA69762 standard; cDNA; 2017 BP.

XX AAA69762;

AC 07-NOV-2000 (first entry)

DE Human ovarian carcinoma antigen polynucleotide SEQ ID NO:72.

XX Human; ovarian carcinoma; ovarian cancer; therapy; diagnosis;

KW tumour antigen; identification; cytosolic; gene therapy; vaccine; ss.

XX Homo sapiens.

OS WO200036107-A2.

XX 22-JUN-2000.

PF 17-DEC-1999; 99WO-US30270.

XX 17-DEC-1998; 98US-0215681.

PR 17-DEC-1998; 98US-0216003.

PR 23-JUN-1999; 99US-0338933.

PR 24-SEP-1999; 99US-0404679.

XX (CORI-) CORIXA CORP.

PI Mltcham JL, King GE, Algate PA, Fridakis TN;

DR WPI; 2000-431589/37.

XX Immunogenic portion of an ovarian carcinoma protein and the nucleic

PT acid encoding it, useful for the diagnosis, prevention and treatment of

PS cancer, preferably ovarian cancer -

XX Claim 1; Fig 2; 299pp; English.

CC The present invention describes an isolated polypeptide comprising an

CC immunogenic portion of an ovarian carcinoma protein (or its variants).

CC Ovarian carcinoma proteins, and polynucleotides encoding them, have

CC cytosolic activity and can be used in gene therapy and vaccines.

CC Ovarian carcinoma polypeptides, nucleic acids, antibodies and vaccines

CC are useful for the prevention, diagnosis and treatment of cancer,

CC preferably ovarian cancer. AAA69691 to AAA70077 and AAB12552 to AAB12557

CC represent human ovarian carcinoma polynucleotides and proteins used in

CC the exemplification of the present invention.

XX Sequence 2017 BP; 673 A; 440 C; 462 G; 442 T; 0 other;

XX Query Match 56.0%; Score 14; DB 21; Length 2017;
 Best Local Similarity 100.0%; Pred. No. 78;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 aggtgcatcgccag 21
 |||||
 Db 964 AGGTGTCATCGCCAG 951

RESULT 16
 AAT30684/c
 ID AAT30684 standard; DNA; 2193 BP.

AC AAT30684;
 XX 11-SEP-1996 (first entry)

DE Kaposi's sarcoma associated herpesvirus ORF22.

KW Kaposi's sarcoma associated herpesvirus; KSHV; gamma-2 herpesvirus;

XX therapy; diagnosis; antisense; triplex; vaccine; AIDS; ss.

OS Kaposi's sarcoma associated herpesvirus.

PN WO9615779-A1.

PD 30-MAY-1996.

PF 21-NOV-1995; 95WO-US15138.

PR 11-APR-1995; 95US-0420235.

PR 21-NOV-1994; 94US-0343101.

PA (UYCO) UNIV COLUMBIA NEW YORK.

PI Chang Y, Moore PS;

DR WPI; 1996-268320/27.

DR P-PSDB; AAR97832.

PT Herpes virus associated with Kaposi's sarcoma - also definitive DNA

PT sequences, useful for diagnosis of and to develop prods. for

PT treatment of Kaposi's sarcoma

PS Claim 17; Page 198-201; 277pp; English.

XX Lambda clone KS5 (AAT30681) has 17 protein-coding regions (AAT30682-98),

CC 15 of which are complete ORFs, including ORF22 (AAT30684) (nts 2033-

CC 4285), which codes for glycoprotein-H (AAR97832). KS5 is a fragment

CC of a newly identified human gamma-2 herpesvirus associated with

CC Kaposi's sarcoma (KS). The ORFs were identified by sequence and

CC positional homology to known herpesvirus sequences. Probes that bind

CC specifically to the KS associated herpesvirus can be used for KS

CC diagnosis. Antisense or triplex oligonucleotides are useful for

CC prophylaxis or treatment of KS, and the protein products (see also

CC AAR97830-46) of the 17 open reading frames are useful as vaccines.

XX Sequence 2193 BP; 510 A; 597 C; 536 G; 550 T; 0 other;

XX Query Match 56.0%; Score 14; DB 17; Length 2193;
 Best Local Similarity 100.0%; Pred. No. 78;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 tgcacgcagcatt 24
 |||||
 Db 46 TGCATCGCCAGCAT 33

DE Kaposi's sarcoma associated herpesvirus ORF22.

XX Kaposi's sarcoma; KSHV; lymphoma; AIDS; vaccine; diagnosis; therapy;

XX glycoprotein H; ss.

OS Kaposi's sarcoma associated herpesvirus.

XX WO9606159-A1.

```

PD 29-FEB-1996.
XX
XX 11-AUG-1995; 95WO-US10194.
XX
XX 11-APR-1995; 95US-0420235.
PR 18-AUG-1994; 94US-0292365.
PR 21-NOV-1994; 94US-0343100.
XX
PA (UNCO ) UNIV COLUMBIA NEW YORK.
PA (GRAN/) GRANT D E.
PA (VIEL/) VIELE L.
XX
PI Chang Y, Moore PS, Grant DE, Viele L;
XX
XX MPI: 1996-151362/15.
DR P-PSDB; AAR93608.
XX
PT Herpesvirus DNA associated with Kaposi's sarcoma - also associated
PT vectors and proteins, used in detection and vaccination.
XX
PS Claim 17; Page 211-214; 305pp; English.
XX
CC Kaposi's sarcoma associated herpes virus (KSHV) clone K35 (AA116806),
CC obtd. from a KS lesion genomic library, includes 15 complete ORFs and
CC 2 incomplete ORFs (AA116807-23) named according to their herpesvirus
CC salmairi positional homologues. ORF22 (AA116814) codes for
CC glycoprotein H (AAR93608). KSHV DNA may be incorporated into a vector
CC and expressed in host cells to produce peptides useful in vaccines or
CC for raising antibodies. The DNA may itself be used to immunise a
CC subject, or used to design therapeutic antisense and triplex
CC molecules or diagnostic probes, or to raise transgenic animals.
XX
SQ Sequence 2193 BP; 510 A; 597 C; 536 G; 550 T; 0 other;

Query Match 56.0%; Score 14; DB 17; Length 2193;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 tgcacgcgcacgat 24
   ||||||||||||
DB 46 TGCATCGCCAGCAT 33

RESULT 18
AAZ20914/C
ID AAZ20914 standard; DNA; 2193 BP.
XX
AC AAZ20914;
XX
DT 02-DEC-1999 (first entry)
XX
DE Nucleotide sequence of gH glycoprotein of HHV-8.
XX
XX vaccine; antibody; Kaposi's sarcoma; glycoprotein; detection;
KW HHV-8; Human Herpesvirus type 8; ds.
XX
OS Human Herpes Virus type 8.
XX
PN MO9945948-A1.
XX
PD 16-SEP-1999.
XX
PF 12-MAR-1999; 99WO-US05464.
XX
PR 12-MAR-1998; 98US-0077670.
XX
PA (UYPE-) UNIV PENNSYLVANIA.
XX
PI Whitbeck JC, Eisenberg RJ, Cohen GH;
XX
XX MPI: 1999-551211/46.
DR P-PSDB; AAY42395.

```

```

XX
XX New nucleic acid molecules encoding human herpesvirus type glycoprotein
PT H and glycoprotein L polypeptides for diagnosis and treatment of
PT Kaposi's Sarcoma
XX
PS Claim 3; Fig 1; 75pp; English.
XX
CC This is the nucleotide sequence of the human herpes virus type 8
CC glycoprotein H. This sequence and it's encoded protein (AAY42395) can be
CC used in the development of vaccines, antibodies and other preparations.
CC The vaccine is useful as a therapeutic agent for prevention of HHV-8
CC disease and as a therapeutic agent for treatment of HHV-8 infection
CC in order to boost the immune response in the animal. Administration of
CC a pure preparation of soluble HHV-8 gH/gL complex or at least one
CC polynucleotide encoding the soluble complex is useful for preventing or
CC treating an HHV-8 infection in a human.
CC Diagnosing of an HHV-8 infection in a human is carried out by contacting
CC a biological sample with an antibody which specifically binds to either
CC glycoprotein or in complex form. If binding of the antibody occurs then
CC the human has HHV-8 infection.
CC Alternatively the infection is diagnosed by addition of one of the
CC HHV-8 glycoproteins or complex and determining whether specific binding
CC of the protein to an antibody in the biological sample occurs. A further
CC diagnostic method involves contacting the biological sample with an
CC HHV-8 glycoprotein H polynucleotide or HHV-8 glycoprotein L
CC polynucleotide and determining whether binding to a nucleic acid in the
CC sample occurs.
XX
SQ Sequence 2193 BP; 510 A; 597 C; 536 G; 550 T; 0 other;

Query Match 56.0%; Score 14; DB 20; Length 2193;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 tgcacgcgcacgat 24
   ||||||||||||
DB 46 TGCATCGCCAGCAT 33

RESULT 19
AAF61014/C
ID AAF61014 standard; DNA; 2397 BP.
XX
AC AAF61014;
XX
DT 16-MAY-2001 (first entry)
XX
DE P. putida KT2440-associated DNA ORF04125.
XX
XX Transgenic plant; detection; probe; amplification; vaccine carrier;
KW microbial production strain; biological remediation; ds.
XX
OS Pseudomonas putida.
XX
PN DE19935088-A1.
XX
PD 01-FEB-2001.
XX
PF 27-JUL-1999; 99DE-1035088.
XX
PR 27-JUL-1999; 99DE-1035088.
XX
PA (TIGR-) TIGR INST GENOMIC RES.
PA (OUTA-) OUTAGEN GMBH.
PA (GBFR-) GES BIOTECHNOLOGISCHE FORSCHUNG MBH.
PA (DKFZ-) DKFZ DEUT KREBSFORSCHUNGSZENTRUM.
PA (MEDI-) MEDIZINISCHE HOCHSCHULE HANNOVER.
XX
XX MPI: 2001-192469/20.
DR
DR New DNA sequences specific for Pseudomonas putida KT2440, useful as
PT safe genetic engineering host, allow detection in presence of other
PT

```

PT related bacteria -
 XX
 PS Claim 1a: Page 62-64; 158pp; German.
 CC This invention describes novel DNA sequences (1) for specific detection
 CC of *Pseudomonas putida* KT2440. The invention also describes (1)
 CC recombinant expression vector containing (1); (2) prokaryotic or
 CC eukaryotic cells transformed or transfected with (1) or the vector of
 CC (1); (3) production of expression products by culturing cells of (2);
 CC (4) expression products, or their fragments, of (1) and synthetic
 CC proteins or peptides with the same sequences (A); (5) poly- or
 CC mono-clonal antibodies (Ab) that react specifically with (A); (6)
 CC hybridoma cells that produce the monoclonal Ab of (5); (7) transgenic
 CC plants that contain transformed or transfected cells of (2); (8)
 CC detecting KT2440 using a labeled (1) or Ab as probe; and (9) DNA chips
 CC carrying one or more (1), (1)', and their fragments, are used as probes
 CC to detect and isolate full-length cDNAs and/or to amplify such cDNAs by
 CC polymerase chain reaction, and for production of transgenic plants. (1),
 CC or antibodies that recognize their expression products, are used for
 CC detecting the presence of KT2440, particularly in presence of other,
 CC even closely related, bacteria. KT2440 is one of the bacteria classified
 CC as safe, by the National Institutes of Health, for genetic engineering
 CC work, e.g. as microbial production strains, for biological remediation
 CC and as vaccine carriers. (1) are exclusive to KT2440 with no significant
 CC homology with sequences in other bacteria (specifically the closely
 CC related pathogen *P. aeruginosa*). Compared with other 'safe' bacteria, it
 CC has greater catabolic activity and better survival in, and adaptation to,
 CC the rhizosphere and soil.
 XX
 SQ Sequence 2397 BP; 488 A; 792 C; 708 G; 406 T; 3 other;
 XX
 Query Match 56.0%; Score 14; DB 22; Length 2397;
 Best Local Similarity 100.0%; Pred. No. 78;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 gcgcagcaggtgca 14
 |||||||
 Db 723 GCGCAGCAGGTGCA 710
 RESULT 20
 AA239011/c
 ID AA239011 standard; cDNA: 3593 BP.
 XX
 AC AA239011:
 XX
 DT 28-FEB-2000 (first entry)
 XX
 DE Mouse Ese2 coding sequence.
 XX
 KW Mouse; murine; Ese1; Ese2; endocytosis; vesicular trafficking;
 KW regulation; actin cytoskeleton; detection; cancer; infection;
 KW EH-domain and SH3-domain regulator of endocytosis; anticancer;
 KW antiproliferative; antiviral; ss.
 XX
 OS Mus sp.
 PN WO9955728-A2.
 PD 04-NOV-1999.
 XX
 PF 27-APR-1999; 99WO-CA00375.
 XX
 PR 27-APR-1998; 98CA-2230201.
 PR 05-FEB-1999; 99US-0118739.
 XX
 PA (HSCR-) HSC RES & DEV LP.
 XX
 PI Egan SE, Wang W, Sengar A;
 XX WPI: 2000-052802/04.
 DR P-PSDB: AAY57445.

XX
 PT New nucleic acid encoding Ese1 and 2 proteins, involved in regulation
 PT of endocytosis, used e.g. for treating cancer or preventing viral
 PT infection -
 PS Claim 25; Page 46-48; 99pp; English.
 XX
 CC The present sequence encodes mouse Ese2. The present invention
 CC specifically describes mammalian Ese1 and 2 proteins (1) and their splice
 CC variants (Ese = EH-domain and SH3-domain regulator of endocytosis). (1)
 CC are involved in regulation of clathrin-mediated endocytosis (as a complex
 CC with Eps15 protein), vesicular trafficking and actin cytoskeleton.
 CC Generally (1) (or its antagonists, mimetics, fragments and inactive
 CC mutants); (1)-specific antibodies (Ab); sequences antisense to the (1)
 CC polynucleotide; agents that downregulate expression of Ese genes or
 CC antagonists of an Ese binding partner are used to treat diseases
 CC associated with undesirable endocytosis and resulting changes in cellular
 CC function. Particularly overexpression of Ese1 is used to block
 CC clathrin-mediated endocytosis in vivo or in cell cultures, while
 CC administration of (1) is used to promote endocytosis of selected cells.
 CC (Ant)agonists of (1) or Ab are used to suppress abnormal proliferation of
 CC cells that can be stimulated to proliferate by a growth factor receptor;
 CC and similar compounds (also inactive Ese mutants) can be used to prevent
 CC viral infection. Endocytosis may also be regulated, in vivo or in cell
 CC cultures, by forming an Ese-Eps15 complex, then binding dynamin to the
 CC complex. Generally conditions that can be treated include cancer;
 CC abnormal cell division or migration; viral infection; or abnormal
 CC receptor signalling, tissue development or synaptic transmission.
 XX
 SQ Sequence 3593 BP; 1171 A; 783 C; 850 G; 789 T; 0 other;
 XX
 Query Match 56.0%; Score 14; DB 21; Length 3593;
 Best Local Similarity 100.0%; Pred. No. 78;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 8 agtgcacgcgcag 21
 |||||||
 Db 923 AGGTGCATCGCCAG 910
 RESULT 21
 AA081778/c
 ID AA081778 standard; DNA: 4182 BP.
 XX
 AC AA081778:
 XX
 DT 09-AUG-1995 (first entry)
 XX
 DE DNA encoding cytohesin protein.
 XX
 KW Cytohesin protein; diagnostic probe; unusual codon usage;
 KW immunosassay; antigen; ds.
 XX
 OS Mycoplasma gallisepticum.
 XX
 Key Location/Qualifiers
 FH 713..4081
 FT CDS /*tag= a
 FT /product= cytohesin protein
 FT misc_difference 887..889
 FT /*tag= b
 FT /codon= seq:tga, aa:trp
 FT misc_difference 998..1000
 FT /*tag= c
 FT /codon= seq:tga, aa:trp
 FT misc_difference 1484..1486
 FT /*tag= d
 FT /codon= seq:cta, aa:lys
 FT misc_difference 1487..1489
 FT /*tag= e
 FT /codon= seq:tta, aa:lys
 FT misc_difference 1559..1561

FT	/tag= f	/codon= seq:tta, aa:lys
FT		
FT	misc_difference	1592..1594
FT	/tag= g	/codon= seq:tta, aa:lys
FT		
FT	misc_difference	1676...1678
FT	/tag= h	/codon= seq:tga, aa:trp
FT		
FT	misc_difference	1715...1717
FT	/tag= i	/codon= seq:ctt, aa:lys
FT		
FT	misc_difference	1853...1855
FT	/tag= j	/codon= seq:tga, aa:trp
FT		
FT	misc_difference	1890...1892
FT	/tag= k	/codon= seq:tga, aa:trp
FT		
FT	misc_difference	1904..1906
FT	/tag= l	/codon= seq:tta, aa:lys
FT		
FT	misc_difference	1961..1963
FT	/tag= m	/codon= seq:tta, aa:lys
FT		
FT	misc_difference	2132...2134
FT	/tag= n	/codon= seq:cta, aa:lys
FT		
FT	misc_difference	2237..2239
FT	/tag= o	/codon= seq:tga, aa:trp
FT		
FT	misc_difference	2273...2275
FT	/tag= p	/codon= seq:tga, aa:trp
FT		
FT	misc_difference	2417..2419
FT	/tag= q	/codon= seq:tga, aa:trp
FT		
FT	misc_difference	2504...2506
FT	/tag= r	/codon= seq:tta, aa:lys
FT		
FT	misc_difference	2609...2611
FT	/tag= s	/codon= seq:tga, aa:trp
FT		
FT	misc_difference	2639...2641
FT	/tag= t	/codon= seq:tga, aa:trp
FT		
FT	misc_difference	2858...2860
FT	/tag= u	/codon= seq:tga, aa:trp
FT		
FT	misc_difference	3020...3022
FT	/tag= v	/codon= seq:tga, aa:trp
FT		
FT	misc_difference	3083...3085
FT	/tag= w	/codon= seq:tga, aa:trp
FT		
FT	misc_difference	3138...3140
FT	/tag= x	/codon= seq:tga, aa:trp
FT		
FT		
PN	US5378820-A.	
PD	03-JAN-1995.	
PR	09-NOV-1992;	92US-0973257.
PX	09-NOV-1992;	92US-0973257.
PA	(DOHM/) DOHMS J. E.	
PA	(KEELR/) KEELER C L.	
PX	Döhms JE, Keeler CL:	
DR	WPI. 1995-051314/07.	
XX	P-PADB; AAR64927.	

PT	Nucleic acid encoding cytohesin protein - used as a probe to
PF	diagnose Mycoplasma gallisepticum infection in poultry
XX	
PS	Claim 1; Column 7-16; 13pp; English.
XX	
CC	The sequence encodes a cytohesin protein from Mycoplasma
CC	gallisepticum. All or part of the sequence may be used as a probe
CC	for diagnosis of M. gallisepticum infection in poultry, e.g. turkey
CC	or fowl. The sequence has several UGA codons, which may be
CC	eliminated to avoid problems of expression in vitro due to irregular
CC	codon usage. The sequence may be cloned in a vector for recombinant
CC	cytohesin production, for use as an antigen in immunoassay formats.
XX	
SQ	Sequence 4182 BP; 1456 A; 763 C; 746 G; 1217 T; 0 other:
OY	3 gcagcaggatgcac 16
Db	1260 GCAGCAGGTGCATC 1247
RESULT 22	
ID	AAZ39010/C
AC	AAZ39010 standard; cDNA; 4625 BP.
XX	
AC	AAZ39010;
XX	
DT	28-FEB-2000 (first entry)
XX	
DE	Mouse Ese2 full length cDNA sequence.
XX	
KM	Mouse; murine; Esel; Ese2; endocytosis; vesicular trafficking;
KM	regulation; actin cytoskeleton; detection; cancer; infection;
KW	Eh-domain and SH3-domain regulator of endocytosis; anticancer;
RW	anti-proliferative; antiviral; ss.
OS	Mus sp.
XX	
PN	WC0955728-A2.
XX	
PD	04-NOV-1999.
XX	
FP	27-APR-1999; 99WO-CA00375.
XX	
PR	27-APR-1998; 98CA-2230201.
PR	05-FEB-1999; 99US-0118739.
PA	(HSCR-) HSC RES & DEV LP.
XX	
PI	Egan SE, Wang W, Sengar A;
XX	
DR	WPI; 2000-052802/04.
XX	
P-PSDB:	AAZY7445.
PT	New nucleic acid encoding Esel and 2 proteins, involved in regulation
PT	of endocytosis, used e.g. for treating cancer or preventing viral
PT	infection -
XX	
PS	Claim 25; Page 43-46; 99pp; English.
XX	
CC	The present sequence encodes mouse Ese2. The present invention
CC	specifically describes mammalian Esel and 2 proteins (I) and their splice
CC	variants (Ese = Eh-domain and SH3-domain regulator of endocytosis). (I)
CC	are involved in regulation of clathrin-mediated endocytosis (as a complex
CC	with Esp1 protein), vesicular trafficking and actin cytoskeleton.
CC	Generally (I) (or its antagonists, mimetics, fragments and inactive
CC	mutants); (II)-specific antibodies (Ab); sequences antisense to the (I)
CC	polynucleotide; agents that downregulate expression of Ese genes or
CC	antagonists of an Ese binding partner are used to treat diseases

CC associated with undesirable endocytosis and resulting changes in cellular
 CC function. Particularly overexpression of Ese1 is used to block
 CC clathrin-mediated endocytosis in vivo or in cell cultures, while
 CC administration of (I) is used to promote endocytosis of selected cells.
 CC (Ant)agonists of (I) or Ab are used to suppress abnormal proliferation of
 CC cells that can be stimulated to proliferate by a growth factor receptor;
 CC and similar compounds (also inactive Ese mutants) can be used to prevent
 CC viral infection. Endocytosis may also be regulated, in vivo or in cell
 CC cultures, by forming an Ese-Esp15 complex, then binding dynamin to the
 CC complex. Generally conditions that can be treated include cancer;
 CC abnormal cell division or migration; viral infection; or abnormal
 CC receptor signaling, tissue development or synaptic transmission.

XX Sequence 4625 BP; 1453 A; 1009 C; 1104 G; 1059 T; 0 other;

Query Match Score 14; DB 21; Length 4625;
 Best Local Similarity 100.0%; Pred. No. 77;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 aggtgcatgcgcag 21
 |||||
 Db 1173 AGGTGCATGCCAG 1160

RESULT 23
 AA39027/C
 ID AA39027 standard; cDNA; 4975 BP.

XX AA39027;

XX 28-FEB-2000 (first entry)

XX Mouse Ese2L coding sequence.

XX Mouse; murine; Ese2; endocytosis; vesicular trafficking;

KW regulation; actin cytoskeleton; detection; cancer; infection;

KW EH-domain and SH3-domain regulator of endocytosis; anticancer;

XX antiproliferative; antiviral; ss.

XX Mus sp.

XX W09955728-A2.

XX 04-NOV-1999.

XX 27-APR-1999; 99WO-CA00375.

XX 27-APR-1998; 98CA-2230201.

XX 05-FEB-1999; 99US-0118739.

XX (HSCR-) HSC RES & DEV LP.

XX Egan SE, Wang W, Sengar A;

XX WPI; 2000-052802/04.

XX P-PSDB; AAY57450.

XX New nucleic acid encoding Ese1 and 2 proteins, involved in regulation

XX of endocytosis, used e.g. for treating cancer or preventing viral

XX infection

XX Claim 25; Page 66-69; 99pp; English.

XX The present invention specifically describes mammalian Ese1 and 2
 CC proteins (I) and their splice variants (Ese = EH-domain and SH3-domain
 CC regulator of endocytosis). (I) are involved in regulation of clathrin-
 CC mediated endocytosis (as a complex with Esp15 protein), vesicular
 CC trafficking and actin cytoskeleton. Generally (I) or its (ant)agonists,
 CC mimetics, fragments and inactive mutants); (I)-specific antibodies (Ab);
 CC sequences antisense to the (I) polynucleotide; agents that downregulate
 CC expression of Ese genes or antagonists of an Ese binding partner are
 CC used to treat diseases associated with undesirable endocytosis and

CC resulting changes in cellular function. Particularly overexpression of
 CC Ese1 is used to block clathrin-mediated endocytosis in vivo or in cell
 CC cultures, while administration of (I) is used to promote endocytosis of
 CC selected cells. (Ant)agonists of (I) or Ab are used to suppress abnormal
 CC proliferation of cells that can be stimulated to proliferate by a growth
 CC factor receptor; and similar compounds (also inactive Ese mutants) can be
 CC used to prevent viral infection. Endocytosis may also be regulated, in
 CC vivo or in cell cultures, by forming an Ese-Esp15 complex, then binding
 CC dynamin to the complex. Generally conditions that can be treated include
 CC cancer; abnormal cell division or migration; viral infection; or abnormal
 CC receptor signaling, tissue development or synaptic transmission. The
 CC present sequence represents mouse Ese2L coding sequence.

XX Sequence 4975 BP; 1540 A; 1166 C; 1201 G; 1068 T; 0 other;

Query Match Score 14; DB 21; Length 4975;
 Best Local Similarity 100.0%; Pred. No. 77;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 aggtgcatgcgcag 21
 |||||
 Db 923 AGGTGCATGCCAG 910

RESULT 24
 AAT97296/C
 ID AAT97296 standard; DNA; 5362 BP.

XX AAT97296;

XX 14-MAY-1998 (first entry)

XX Mouse additional sex comb polypeptide encoding DNA.

KW Mouse; additional sex comb; Asx; cancer; anti-proliferative;

KW dysregulation; de-differentiation; oncogene; ss.

XX Mus sp.

XX W09742213-A1.

XX 13-NOV-1997.

XX 08-MAY-1997; 97WO-US07806.

XX 08-JUL-1996; 96US-0021383.

XX 08-MAY-1996; 96US-0016585.

XX (CHIR) CHIRON CORP.

XX Randazzo F;

XX WPI; 1997-558903/51.

XX P-PSDB; AAW37392.

XX New mammalian additional sex comb polypeptide(s) and related nucleic

XX acid - used for diagnosis of cancer, screening for

XX anti-proliferative compounds and for inducing de-differentiation and

XX dysregulation of cells

XX Example 5; Page 44-48; 64pp; English.

XX The present sequence encodes mouse Asx (additional sex comb)
 CC polypeptide. Overexpression of Asx, or expression of a variant Asx, is
 CC indicative of cancer, especially colorectal adenocarcinoma, carcinoma,
 CC melanoma, lymphoma or leukemia. Diagnosis (or monitoring of disease) is
 CC done using DNA encoding Asx-based probes; amplification with DNA

CC encoding Asx-derived primers; hybridisation with branched DNA (bDNA)
 CC probes that are specific for DNA encoding Asx; using Ab or by
 CC quantitation of RNA. Modulators that decrease transcription of Asx are
 CC potentially useful as antineoplastic agents, e.g. for treating cancers
 CC specified above or other proliferative disorders such as restenosis,
 CC benign prostatic hypertrophy, retinopathy, psoriasis, arthritis and
 CC premalignant lesions. These conditions can also be treated with DNA
 CC encoding Asx in standard gene therapy processes. The Asx gene promoter
 CC is useful for tissue-specific expression (e.g. in treatment of lung or
 CC colorectal cancers) and the untranslated regions can be used as probes.
 CC Also Asx (1) induces de-differentiation by treatment of progenitor cells
 CC and (ii) causes dysregulation of controlled cell growth. Further it is
 CC used to raise Ab (useful as diagnostic immunoassay reagents or
 CC therapeutically for inhibiting DNA encoding Asx activity); as competitor
 CC in immunoassays and for purification of Ab.
 CC
 SQ Sequence 5362 BP; 1430 A; 1354 C; 1399 G; 1179 T; 0 other;

Query Match 56.0%; Score 14; DB 18; Length 5362;
 Best Local Similarity 100.0%; Pred. No. 77;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 gcacgagtgatc 16
 |||||
 DB 3354 GCACGAGTGATC 3341

RESULT 25
 AAT97294/C
 ID AAT97294 standard; DNA: 5362 BP.
 XX
 XX AAT97294;

DT 14-MAY-1998 (first entry)

DE Mouse additional sex comb polypeptide encoding DNA.

KW Mouse; additional sex comb; Asx: tumour suppressor; anti-proliferative;
 KW cancer; differentiation; cell growth; ss.

OS Mus sp.

FX Key Location/Qualifiers
 FT CDS 93..4637
 FT /*tag= a
 FT /product= "Asx"

PN W09742219-A1.

PD 13-NOV-1997.

PF 08-MAY-1997; 97MO-US07805.

PR 08-JUL-1996; 96US-0021383.

PR 08-MAY-1996; 96US-0016585.

PA (CHIR) CHIRON CORP.

PI Randazzo F;

DR WPI: 1997-558906/51.

DR P-PSDB; AAM37390.

PT New mammalian additional sex comb proteins and related nucleic acid
 PT - used for diagnosis and treatment of cancer, screening for
 PT anti-proliferative compounds and for inducing cell growth and
 PT differentiation

PS Example 5; Page 39-42; 55pp; English.

CC The present sequence encodes mouse Asx (additional sex comb) protein.
 CC Over- or under-expression of Asx, or expression of a variant Asx, is

CC indicative of cancer, especially colorectal adenocarcinoma, melanoma,
 CC lymphoma or leukaemia. Diagnosis (or monitoring of disease) is done
 CC using DNA encoding Asx-based probes; amplification with DNA encoding
 CC Asx-derived primers; hybridisation with branched DNA (bDNA) probes that
 CC are specific for DNA encoding Asx; using Ab or by quantitation of RNA.
 CC Modulators that increase transcription of Asx are potentially useful as
 CC antineoplastic agents, e.g. for treating cancers specified above or
 CC other proliferative disorders such as restenosis, benign prostatic
 CC hypertrophy, retinopathy, psoriasis, arthritis and premalignant lesions.
 CC These conditions can also be treated with DNA encoding Asx in standard
 CC gene therapy processes. The Asx gene promoter is useful for tissue-
 CC specific expression (e.g. in treatment of lung or colorectal cancers)
 CC and the untranslated regions can be used as probes. Also Asx is used (1)
 CC to treat neoplasia; (ii) to induce differentiation by treatment of
 CC progenitor cells and (iii) to regulate growth in cells where growth is
 CC dysregulated. Further it is used to raise Ab (useful as diagnostic
 CC immunoassay reagents or therapeutically for inhibiting Asx activity); as
 CC competitor in immunoassays and for purification of Ab.
 CC
 SQ Sequence 5362 BP; 1430 A; 1354 C; 1399 G; 1179 T; 0 other;

Query Match 56.0%; Score 14; DB 18; Length 5362;
 Best Local Similarity 100.0%; Pred. No. 77;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 gcacgagtgatc 16
 |||||
 DB 3354 GCACGAGTGATC 3341

RESULT 26
 AAV21186/C
 ID AAV21186 standard; DNA: 5676 BP.
 XX
 XX AAV21186;

DT 24-JUL-1998 (first entry)

DE Amycolatopsis mediterranei strain wt3136 5.7 kb KpnI DNA fragment.

KW Amycolatopsis mediterranei; rifamycin; synthesis; gene cluster;
 KW polyketide synthase; actinomycete; ansamycin; ds.

OS Amycolatopsis mediterranei.

FX Key Location/Qualifiers
 FT CDS 3..5676
 FT /*tag= a
 FT /product= "polyketide synthase"
 FT /note= "no stop codon given"

PN W09807868-A1.

PD 26-FEB-1998.

PF 18-AUG-1997; 97MO-EP04495.

PR 20-AUG-1996; 96EP-0810551.

PA (NOVS) NOVARTIS AG.

PI Engel N, Schupp T, Toupet C;

DR WPI: 1998-169172/15.

DR P-PSDB; AAM52844.

PT Amycolatopsis mediterranei rifamycin synthesis gene cluster - used
 PT to produce rifamycin and rifamycin analogues

PS Claim 4; Page 38-43; 205pp; English.

CC The present sequence represents Amycolatopsis mediterranei strain wt3136

CC 5.7 kb KpnI fragment DNA, from the present invention. The present
CC invention describes a Amycolatopsis mediterranei rifamycin synthesis
CC gene cluster DNA fragment comprising a DNA region involved directly or
CC indirectly in the gene cluster responsible for rifamycin synthesis,
CC including the adjacent DNA regions to the right and left which, by
CC reason of their function in connection with rifamycin biosynthesis,
CC qualify as constituents of this rifamycin gene cluster, and functional
CC fragments, derivatives or constituents of these. The Amycolatopsis
CC mediterranei rifamycin synthesis gene cluster DNA fragment can be used
CC for producing rifamycin, rifamycin analogues or precursors. It can also
CC be used for inactivating or modifying genes involved in ansamycin or
CC rifamycin biosynthesis. The DNA can be used for constructing mutant
CC actinomycetes strains from which the natural rifamycin or ansamycin
CC biosynthesis gene cluster has been partly or completely deleted. The
CC DNA fragment can be used for assembling a library of polypeptides
CC synthases, which can be used for assembling a library of polypeptides.
CC A hybridisation probe of the invention can be used for identifying DNA
CC fragments involved in the biosynthesis of ansamycins.

XX
SQ Sequence 5676 BP; 674 A; 1990 C; 2205 G; 807 T; 0 other;

Query Match 56.0%; Score 14; DB 19; Length 5676;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 caggtgcacgcga 20
|||||
DB 3923 CAGGTGCATCGCCA 3910

RESULT 27
AAZ39026/C
ID AAZ39026 standard; cDNA; 6014 BP.

XX
AC AAZ39026;
XX
DT 28-FEB-2000 (first entry)
XX
DE Mouse Ese2L cDNA sequence.
XX
KW Mouse; murine; Ese1; Ese2; endocytosis; vesicular trafficking;
KW regulation; actin cytoskeleton; detection; cancer; infection;
KW EH-domain and SH3-domain regulator of endocytosis; anticancer;
KW antiproliferative; antiviral; ss.
XX
OS Mus sp.
XX
PN W09955728-A2.
XX
PD 04-NOV-1999.
XX
PF 27-APR-1999; 99WO-CA00375.
XX
PR 27-APR-1998; 98CA-2230201.
PR 05-FEB-1999; 99US-0118739.
XX
PA (HSCR-) HSC RES & DEV LP.
XX
PI Egan SE, Wang W, Sengar A;
XX
DR WPI: 2000-052802/04.
DR P-PSDB; AAY57450.
XX
PT New nucleic acid encoding Ese1 and 2 proteins, involved in regulation
PT of endocytosis, used e.g. for treating cancer or preventing viral
PT infection -
XX
PS Claim 25; Page 63-66; 99pp; English.
XX
CC The present invention specifically describes mammalian Ese1 and 2
CC proteins (I) and their splice variants (Ese - EH-domain and SH3-domain
CC regulator of endocytosis). (I) are involved in regulation of clathrin-

CC mediated endocytosis (as a complex with Esp15 protein), vesicular
CC trafficking and actin cytoskeleton. Generally (I) (or its (ant)agonists,
CC mimetics, fragments and inactive mutants); (I)-specific antibodies (Ab);
CC sequences antisense to the (I) polynucleotide; agents that downregulate
CC expression of Ese genes or antagonists of an Ese binding partner are
CC used to treat diseases associated with undesirable endocytosis and
CC resulting changes in cellular function. Particularly overexpression of
CC Ese1 is used to block clathrin-mediated endocytosis in vivo or in cell
CC cultures, while administration of (I) is used to promote endocytosis of
CC selected cells. (Ant)agonists of (I) or Ab are used to suppress abnormal
CC proliferation of cells that can be stimulated to proliferate by a growth
CC factor receptor; and similar compounds (also inactive Ese mutants) can be
CC used to prevent viral infection. Endocytosis may also be regulated, in
CC vivo or in cell cultures, by forming an Ese-Esp15 complex, then binding
CC dynamin to the complex. Generally conditions that can be treated include
CC cancer; abnormal cell division or migration; viral infection; or abnormal
CC receptor signalling, tissue development or synaptic transmission. The
CC present sequence represents mouse Ese2L cDNA sequence.

XX
SQ Sequence 6014 BP; 1833 A; 1389 C; 1454 G; 1338 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 6014;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 aggtgcacgcgag 21
|||||
DB 1173 AGGTGCATCGCCAG 1160

RESULT 28
AAA81733/C
ID AAA81733 standard; DNA; 12893 BP.

XX
AC AAA81733;
XX
DT 04-DEC-2000 (first entry)
XX
DE N. meningitidis partial DNA sequence gmm_280 SEQ ID NO:280.
XX
KW Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;
KW antigen; vaccine; diagnosis; infection; antibacterial; identification;
KW Meningococcus B; MenB; ds.
XX
OS Neisseria meningitidis.
XX
PN W0200022430-A2.
XX
PD 20-APR-2000.
XX
PF 08-OCT-1999; 99WO-US23573.
XX
PR 09-OCT-1998; 98US-0103794.
PR 30-APR-1999; 99US-0132068.
XX
PA (CHIR) CHIRON CORP.
XX
PI Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;
PI Maignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;
PI Rapunoli R, Pizzi M;
XX
DR WPI: 2000-318079/27.
XX
PT Isolated nucleotide sequences of Neisseria meningitidis which can be
PT used in the diagnosis and treatment of N. meningitidis infection and
PT other Neisserial infections, for example, N.gonorrhoea -
XX
PS Claim 7; Page 1561-1565; 1760pp; English.
XX
CC The present invention describes methods of obtaining immunogenic
CC proteins from Neisseria genomic sequences. AAA81453 to AAA82414
CC represent specifically claimed Neisseria meningitidis genomic DNA

CC sequences: AAA81260 to AAA81303 and AAB25620 to AAB25663 represent
 CC Neisseria DNA sequences and their corresponding proteins; AAA81254 to
 CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
 CC isolation of Neisseria meningitidis DNA sequences; and AAA81322 to
 CC AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF
 CC sequences, which are all used in the exemplification of the present
 CC invention. The nucleic acid sequences, protein sequences, and antibodies
 CC against them, can be used in the manufacture of a composition. The
 CC composition can be used as a medicament (or in the manufacture of a
 CC medicament) for treating, preventing or diagnosing infection due to
 CC Neisseria bacteria. For example, some of the identified proteins could
 CC be components of vaccines against Meningococcus B; against all serotypes;
 CC and/or against all pathogenic Neisseriae. Identification of sequences
 CC from the bacterium will also facilitate production of biological probes,
 CC particularly organism-specific probes. Attempts to make efficacious
 CC meningococcus B vaccines have failed mainly due to antigen tolerance.
 CC Multivalent vaccines have also been tried but none have successfully
 CC overcome antigenic variability. The provision of further, complete
 CC sequences may provide an opportunity to identify secreted or surface
 CC exposed proteins that may be presumed targets for the immune system and
 CC which are not antigenically variable or at least more conserved than
 CC other more variable regions.

XX
 SQ Sequence 12893 BP; 2829 A; 3164 C; 3741 G; 3159 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 12893;
 Best Local Similarity 100.0%; Pred. No. 76;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ggcagcagtgatc 14
 |||||
 Db 3743 GCGCAGCAGTGCA 3730

RESULT 29
 AAV62159
 ID AAV62159 standard; DNA; 20387 BP.
 XX
 AC AAV62159;
 XX
 DT 18-DEC-1998 (first entry)
 XX
 DE HSV-2 strain SB5 Contig ID 15 DNA sequence.
 XX
 KM HSV-2 strain SB5; immunological response induction; therapy;
 KW antiviral identification; viral protein inhibitor; ss.
 XX
 OS Herpes simplex virus type 2.
 FH
 FH Key Location/Qualifiers
 FT complement (1..772)
 FT /tag- a
 FT /transl_except= (pos: 1, aa: Xaa-Xaa-Xaa-Xaa-Xaa)
 FT /product= "ORF#1 protein"
 FT /note= "encoded protein shown in AAW72114;
 FT Xaa = unspecified amino acid"
 FT 802..2073
 FT /tag- b
 FT /product= "ORF#2 protein"
 FT /note= "encoded protein shown in AAW72115"
 FT complement (2413..3504)
 FT /tag- c
 FT /product= "ORF#3 protein"
 FT /note= "encoded protein shown in AAW72116"
 FT complement (3748..7938)
 FT /tag- d
 FT /product= "ORF#4 protein"
 FT /note= "encoded protein shown in AAW72117"
 FT complement (8185..8853)
 FT /tag- e
 FT /product= "ORF#5 protein"
 FT /note= "encoded protein shown in AAW72118"

FT CDS 9189..11102
 FT /tag- f
 FT /product= "ORF#6 protein"
 FT /note= "encoded protein shown in AAW72119"
 FT complement (11323..14091)
 FT /tag- g
 FT /product= "ORF#7 protein"
 FT /note= "encoded protein shown in AAW72120"
 FT complement (14176..15306)
 FT /tag- h
 FT /product= "ORF#8 protein"
 FT /note= "encoded protein shown in AAW72121"
 FT 14989..16053
 FT /tag- i
 FT /product= "ORF#9 protein"
 FT /note= "encoded protein shown in AAW72122"
 FT 16334..18091
 FT /tag- j
 FT /product= "ORF#10 protein"
 FT /note= "encoded protein shown in AAW72123"
 FT 18312..20240
 FT /tag- k
 FT /product= "ORF#11 protein"
 FT /note= "encoded protein shown in AAW72124"

XX WO9820016-A1.
 XX 14-MAY-1998.
 XX
 XX 31-OCT-1997; 97WO-US20016.
 XX
 XX 09-JUN-1997; 97US-0049018.
 XX 04-NOV-1996; 96DS-0030279.
 XX
 XX (SMIK) SMITHKLINE BEECHAM CORP.

XX Chan JY, Dabrowski-Amarel CE, Delvecchio AM, Dillon SB;
 PI Esser KM, Leary JG;
 XX
 XX WPI; 1998-286847/25.
 DR P-PSDB; AAW72114, AAW72115, AAW72116, AAW72117, AAW72118, AAW72119,
 DR AAW72120, AAW72121, AAW72122, AAW72123, AAW72124.
 XX

PT Herpes simplex virus type-2 sequences - useful in, e.g. prevention
 PT and treatment of infection or inducing immunological response in
 PT mammal

XX Claim 1; Page 410-418; 748pp; English.

XX This sequence represents a Herpes simplex virus type-2 (HSV-2) DNA
 CC sequence of the invention. This sequence was isolated from HSV-2 strain
 CC SB5 (deposited as ATCC VR-2546), is designated Contig ID 15, and encodes
 CC 11 HSV-2 proteins. The proteins can be used for the treatment or
 CC prevention of disease, to induce an immunological response in a mammal or
 CC to identify inhibitors, activators or novel antivirals. Antagonists of
 CC the proteins can be used to inhibit a viral polypeptide. The DNA sequence
 CC or a vector containing it can also be used to induce an immunological
 CC response in a mammal.

SQ Sequence 20387 BP; 3318 A; 6905 C; 7034 G; 3130 T; 0 other;

Query Match 56.0%; Score 14; DB 19; Length 20387;
 Best Local Similarity 100.0%; Pred. No. 75;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 ggcagcagtgatc 16
 |||||
 Db 16597 ggcagcagtgatc 16610

RESULT 30
 AAT30681/C

```

ID  AAT30681 standard; DNA; 20710 BP.
XX
AC  AAT30681;
XX
DT  11-SEP-1996 (first entry)
XX
DE  Kaposi's sarcoma associated herpesvirus clone KS5.
XX
KM  Kaposi's sarcoma associated herpesvirus; KSHV; gamma-2 herpesvirus;
KW  therapy; diagnosis; antisense; triplex; vaccine; AIDS; ss.
XX
OS  Kaposi's sarcoma associated herpesvirus.
XX
FH  Location/Qualifiers
FT  complement (1..364)
CDS   /tag= a
      /note= "incomplete ORF20"
      363..2105
      /tag= b
      /product= thymidine kinase
      /note= "ORF21"
      2093..4285
      /tag= c
      /product= glycoprotein H
      /note= "ORF22"
      complement (4282..5496)
      /tag= d
      /note= "ORF23"
      complement (5500..7758)
      /tag= e
      /note= "ORF24"
      7757..11887
      /tag= f
      /note= "ORF25"
      11913..12830
      /tag= g
      /product= virion polypeptide VP23
      /note= "ORF26"
      12851..13723
      /tag= h
      /product= minor capsid protein
      /note= "ORF27"
      13969..14331
      /tag= i
      /note= "ORF28"
      complement (14343..15485)
      /tag= j
      /product= (with ORF29A) putative DNA packaging
      protein
      /note= "ORF29B"
      15609..15842
      /tag= k
      /note= "ORF30"
      15839..16423
      /tag= l
      /note= "ORF31"
      16390..17754
      /tag= m
      /note= "ORF32"
      17747..18685
      /tag= n
      /note= "ORF33"
      complement (18743..19662)
      /tag= o
      /product= (with ORF29B) putative DNA packaging
      protein
      /note= "ORF29A"
      19661..20644
      /tag= p
      /note= "ORF34"
      20625..20710
      /tag= q
      /note= "incomplete ORF35"
      20710

```

```

XX  XX  W09615779-A1.
XX  XX  30-MAY-1996.
XX  XX  21-NOV-1995; 95WO-US15138.
XX  XX  11-APR-1995; 95US-0420235.
XX  XX  21-NOV-1994; 94US-0343101.
XX  PA  (UYCO ) UNIV COLUMBIA NEW YORK.
XX  PI  Chang Y, Moore PS;
XX  DR  WPI: 1996-268320/27.
XX  PT  Herpes virus associated with Kaposi's sarcoma - also definitive DNA
XX  PT  sequences, useful for diagnosis of and to develop prods. for
XX  PT  treatment of Kaposi's sarcoma
XX  PS  Claim 16; Page 164-175; 277pp; English.
XX  CC  Phage lambda clone KS5 (AAT30681) was isolated from an AIDS-Kaposi's
XX  CC  sarcoma (KS) genomic DNA library by screening with a specific DNA,
XX  CC  KS330Bam, found in KS lesions. Sequence analysis of KS5 showed
XX  CC  that KS330Bam is part of a larger herpesvirus genome. This agent
XX  CC  was defined as a human gamma-2 herpesvirus that can be continuously
XX  CC  cultured in naturally-transformed, EBV-coinfected lymphocytes from
XX  CC  AIDS-related body-cavity based lymphomas. KS5 has 17 protein-coding
XX  CC  regions (see also AAT30682-98) identified by homology to known gamma-
XX  CC  herpesvirus sequences. Probes that bind specifically to the KS
XX  CC  associated herpesvirus can be used for KS diagnosis. Antisense or
XX  CC  triplex oligonucleotides are useful for prophylaxis or treatment of
XX  CC  KS, and the peptide products (see also AA97830-46) of the open reading
XX  CC  frames are useful as vaccines.
XX  SQ  Sequence 20710 BP; 4628 A; 5526 C; 5666 G; 4890 T; 0 other;

Query Match 56.0%; Score 14; DB 17; Length 20710;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy  11  tgcacgcacacat 24
    |||||
Db  2138  TGCATGCCACGCAAT 2125

RESULT 31
AAT16806/c
ID  AAT16806 standard; DNA; 20710 BP.
XX
AC  AAT16806;
XX
DT  13-AUG-1996 (first entry)
XX
DE  Kaposi's sarcoma associated herpesvirus clone KS5.
XX
KM  Kaposi's sarcoma; KSHV; lymphoma; AIDS; vaccine; diagnosis; therapy;
KW  ss.
XX
OS  Kaposi's sarcoma associated herpesvirus.
XX
PA  W09606159-A1.
XX
PI  29-FEB-1996.
XX
PD  11-AUG-1995; 95WO-US10194.
XX
PE  11-AUG-1995; 95US-0420235.
XX
PR  11-APR-1995; 95US-0420235.
XX
PR  18-AUG-1994; 94US-0292365.
XX
PR  21-NOV-1994; 94US-0343100.
XX

```

PA (UYCO) UNIV COLUMBIA NEW YORK.
 PA (GRAN/) GRANT D E.
 PA (VIEL/) VIELE L.
 XX
 PI Chang Y, Moore PS, Grant DE, Viele L;
 XX
 DR WPI: 1996-151362/15.
 XX
 PT Herpesvirus DNA associated with Kaposi's sarcoma - also associated
 PT vectors and proteins, used in detection and vaccination.
 XX
 PS Claim 16, Page 177-188; 305pp; English.
 XX
 CC Representational difference analysis was used to identify unique
 CC DNA sequences in Kaposi's sarcoma (KS) tissue. An isolated sequence,
 CC K5330bam (AA16824), was used to screen a KS lesion genomic library. A
 CC lambda phage, K55, was identified and the sequence of its insert
 CC (AA16808) was determined. This represents a novel infectious agent
 CC associated with AIDS-KS, Kaposi's sarcoma associated herpesvirus (KSHV).
 CC The K55 fragment has 15 complete ORFs and 2 incomplete ORFs (see also
 CC AA16807-23). Portions of the sequence may be incorporated into a
 CC vector and expressed in host cells to produce peptides (see also
 CC AAR93601-17) useful in vaccines or for raising antibodies. The DNA may
 CC itself be used to immunise a subject, or used to design therapeutic
 CC antisense and triplex molecules or diagnostic probes, or to raise
 CC transgenic animals.
 XX
 S0 Sequence 20710 BP; 4628 A; 5526 C; 5666 G; 4890 T; 0 other;
 Query Match 56.0%; Score 14; DB 17; Length 20710;
 Best Local Similarity 100.0%; Pred. No. 75;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 11 tgcacgcgcacgat 24
 ||||||||||||
 DB 2138 TGCATCGCCACGAT 2125
 RESULT 32
 AAV62134
 ID AAV62134 standard; DNA; 26338 BP.
 XX
 AC AAV62134;
 XX
 DT 07-DEC-1998 (first entry)
 XX
 DE HSV-2 strain SB5 Contig ID 104 DNA sequence.
 XX
 KW HSV-2 strain SB5; immunological response induction; therapy;
 KW antiviral identification; viral protein inhibitor; ss.
 XX
 OS Herpes simplex virus type 2.
 XX
 FH Location/Qualifiers
 FH complement (256..3021)
 FT /*tag= a
 FT /product= "ORF#1 protein"
 FT /note= "encoded protein shown in AAW72033"
 FT complement (2895..3908)
 FT /*tag= b
 FT /product= "ORF#2 protein"
 FT /note= "encoded protein shown in AAW72034"
 FT complement (3637..6093)
 FT /*tag= c
 FT /product= "ORF#3 protein"
 FT /note= "encoded protein shown in AAW72035"
 FT complement (6263..6745)
 FT /*tag= d
 FT /product= "ORF#4 protein"
 FT /note= "encoded protein shown in AAW72036"
 FT 6775..8046
 FT /*tag= e

FT /product= "ORF#5 protein"
 FT /note= "encoded protein shown in AAW72037"
 FT complement (8386..9477)
 FT /*tag= f
 FT /product= "ORF#6 protein"
 FT /note= "encoded protein shown in AAW72038"
 FT complement (9721..13911)
 FT /*tag= g
 FT /product= "ORF#7 protein"
 FT /note= "encoded protein shown in AAW72039"
 FT complement (14158..14826)
 FT /*tag= h
 FT /product= "ORF#8 protein"
 FT /note= "encoded protein shown in AAW72040"
 FT 15162..17075
 FT /*tag= i
 FT /product= "ORF#9 protein"
 FT /note= "encoded protein shown in AAW72041"
 FT complement (17295..19110)
 FT /*tag= j
 FT /product= "ORF#10 protein"
 FT /note= "encoded protein shown in AAW72042"
 FT complement (19111..20064)
 FT /*tag= k
 FT /product= "ORF#11 protein"
 FT /note= "encoded protein shown in AAW72043"
 FT complement (20149..21279)
 FT /*tag= l
 FT /product= "ORF#12 protein"
 FT /note= "encoded protein shown in AAW72044"
 FT 20962..22026
 FT /*tag= m
 FT /product= "ORF#13 protein"
 FT /note= "encoded protein shown in AAW72045"
 FT 22307..23887
 FT /*tag= n
 FT /product= "ORF#14 protein"
 FT /note= "encoded protein shown in AAW72046"
 FT 23778..24065
 FT /*tag= o
 FT /product= "ORF#15 protein"
 FT /note= "encoded protein shown in AAW72047"
 FT 24286..25632
 FT /*tag= p
 FT /product= "ORF#16 protein"
 FT /note= "encoded protein shown in AAW72048"
 FT 25457..26215
 FT /*tag= q
 FT /product= "ORF#17 protein"
 FT /note= "encoded protein shown in AAW72049"
 XX
 PN MO9820016-A1.
 XX
 PD 14-MAY-1998.
 XX
 PF 31-OCT-1997; 97WO-US20016.
 XX
 PR 09-JUN-1997; 97US-0049018.
 PR 04-NOV-1996; 96US-0030279.
 XX
 PA (SMK) SMITHKLINE BEECHAM CORP.
 XX
 PI Chan JY, Dabrowski-Amara CE, Delvecchio AM, Dillon SB;
 PI Esser KM, Leary JJ;
 XX
 DR WPI: 1998-286847/25.
 DR P-RSDB: AAW72033, AAW72034, AAW72035, AAW72036, AAW72037, AAW72038,
 DR AAW72039, AAW72040, AAW72041, AAW72042, AAW72043, AAW72044, AAW72045,
 DR AAW72046, AAW72047, AAW72048, AAW72049.
 XX
 PT Herpes simplex virus type-2 sequences - useful in, e.g. prevention
 PT and treatment of infection or inducing immunological response in
 PT mammal

PK Z0-AUG-1990; 90EP-0810551.
XY

PA (NOVS) NOVARTIS AG.
 XX
 PI Engel N, Schupp T, Toupet C;
 XX
 DR WPI: 1998-169172/15,
 DR P-PSDB; AAW52845-W52850.
 XX
 PT Amycolatopsis mediterranei rifamycin synthesis gene cluster - used
 PT to produce rifamycin and rifamycin analogues
 XX
 PS Claim 4; Page 53-102; 205pp; English.
 XX
 CC The present sequence represents a Amycolatopsis mediterranei rifamycin
 CC synthesis gene cluster DNA fragment from the present invention. The
 CC DNA fragment comprises a DNA region involved directly or indirectly
 CC in the gene cluster responsible for rifamycin synthesis, including
 CC the adjacent DNA regions to the right and left which, by reason of
 CC their function in connection with rifamycin biosynthesis, qualify
 CC as constituents of this rifamycin gene cluster, and functional
 CC mediterranei rifamycin synthesis gene cluster DNA fragment can be used
 CC for producing rifamycin, rifamycin analogues or precursors. It can also
 CC be used for inactivating or modifying genes involved in ansamycin or
 CC rifamycin biosynthesis. The DNA can be used for constructing mutant
 CC actinomycetes strains from which the natural rifamycin or ansamycin
 CC biosynthesis gene cluster has been partly or completely deleted. The
 CC DNA fragment can be used for assembling a library of polypeptides.
 CC synthases, which can be used for assembling a library of polypeptides.
 CC A hybridisation probe of the invention can be used for identifying DNA
 CC fragments involved in the biosynthesis of ansamycins.
 XX
 SQ Sequence 53789 BP; 6707 A; 19183 C; 20504 G; 7395 T; 0 other;
 Query Match 56.0%; Score 14; DB 19; Length 53789;
 Best Local Similarity 100.0%; Pred. No. 74;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 7 caggtgcacgcaca 20
 |||||
 Db 47015 CAGGTGCATCGCCA 47002
 RESULT 35
 AAV62176
 ID AAV62176 standard; DNA; 117213 BP.
 XX
 AC AAV62176;
 XX
 DT 13-JAN-1999 (first entry)
 XX
 DE HSV-2 strain SB5 Contig ID 15 DNA sequence.
 XX
 KW HSV-2 strain SB5; immunological response induction; therapy;
 KW antiviral identification; viral protein inhibitor; ss.
 XX
 OS Herpes simplex virus type 2.
 XX
 FH Key
 FT Location/Qualifiers
 FT 755..1297
 FT CDS
 FT /tag= a
 FT /product= "ORF#1 protein"
 FT /note= "encoded protein shown in AAW72170"
 FT 1170..2174
 FT CDS
 FT /tag= b
 FT /product= "ORF#2 protein"
 FT /note= "encoded protein shown in AAW72171"
 FT 2229..2930
 FT CDS
 FT /tag= c
 FT /product= "ORF#3 protein"
 FT /note= "encoded protein shown in AAW72172"
 FT complement (3130..3735)
 FT /tag= q
 FT CDS

FT /product= "ORF#4 protein"
 FT /note= "encoded protein shown in AAW72173"
 FT complement (3802..6447)
 FT CDS
 FT /tag= e
 FT /product= "ORF#5 protein"
 FT /note= "encoded protein shown in AAW72174"
 FT 6017..8482
 FT CDS
 FT /tag= f
 FT /product= "ORF#6g protein"
 FT /note= "encoded protein shown in AAW72250"
 FT 6026..8482
 FT CDS
 FT /tag= g
 FT /product= "ORF#6f protein"
 FT /note= "encoded protein shown in AAW72249"
 FT 6065..8482
 FT CDS
 FT /tag= h
 FT /product= "ORF#6e protein"
 FT /note= "encoded protein shown in AAW72248"
 FT 6167..8482
 FT CDS
 FT /tag= i
 FT /product= "ORF#6d protein"
 FT /note= "encoded protein shown in AAW72247"
 FT 6296..8482
 FT CDS
 FT /tag= j
 FT /product= "ORF#6c protein"
 FT /note= "encoded protein shown in AAW72246"
 FT 6326..8482
 FT CDS
 FT /tag= k
 FT /product= "ORF#6b protein"
 FT /note= "encoded protein shown in AAW72245"
 FT 6446..8482
 FT CDS
 FT /tag= l
 FT /product= "ORF#6a protein"
 FT /transl_except= (pos: 7400..7402, aa: Ala-Ala)
 FT /transl_except= (pos: 7481..7486, aa: Ile)
 FT /note= "encoded protein shown in AAW72244"
 FT 8457..9347
 FT CDS
 FT /tag= m
 FT /product= "ORF#7 protein"
 FT /note= "encoded protein shown in AAW72175"
 FT complement (9604..11855)
 FT CDS
 FT /tag= n
 FT /transl_except= (pos: 11635..11636, aa: Ala)
 FT /product= "ORF#8 protein"
 FT /note= "encoded protein shown in AAW72176"
 FT complement (11905..14508)
 FT CDS
 FT /tag= o
 FT /product= "ORF#9b protein"
 FT /note= "encoded protein shown in AAW72222"
 FT complement (11905..14520)
 FT CDS
 FT /tag= p
 FT /product= "ORF#9a protein"
 FT /note= "encoded protein shown in AAW72223"
 FT 14399..15802
 FT CDS
 FT /tag= q
 FT /product= "ORF#10 protein"
 FT /note= "encoded protein shown in AAW72177"
 FT complement (15996..16286)
 FT CDS
 FT /tag= r
 FT /product= "ORF#11 protein"
 FT /note= "encoded protein shown in AAW72178"
 FT complement (16202..18064)
 FT CDS
 FT /tag= s
 FT /product= "ORF#12 protein"
 FT /note= "encoded protein shown in AAW72179"
 FT complement (18105..19661)
 FT CDS
 FT /tag= t
 FT /product= "ORF#13 protein"
 FT /note= "encoded protein shown in AAW72180"
 FT complement (19415..20074)
 FT CDS
 FT /tag= u
 FT /product= "ORF#14 protein"
 FT /note= "encoded protein shown in AAW72181"
 FT


```

FT CDS 20155..21453
FT /tag= v
FT /product= "ORF#15 protein"
FT /note= "encoded protein shown in AAW72182"
FT CDS complement (21326..22291)
FT /tag= w
FT /product= "ORF#16 protein"
FT /note= "encoded protein shown in AAW72183"
FT CDS complement (22546..24654)
FT /tag= x
FT /product= "ORF#17 protein"
FT /note= "encoded protein shown in AAW72184"
FT CDS 24684..25955
FT /tag= y
FT /product= "ORF#18 protein"
FT /note= "encoded protein shown in AAW72185"
FT CDS complement (26295..27251)
FT /tag= z
FT /product= "ORF#19 protein"
FT /note= "encoded protein shown in AAW72186"
FT CDS complement (27630..31754)
FT /tag= aa
FT /product= "ORF#20b protein"
FT /note= "encoded protein shown in AAW72225"
FT CDS complement (27630..31784)
FT /tag= ab
FT /product= "ORF#20a protein"
FT /note= "encoded protein shown in AAW72224"
FT CDS complement (32067..32735)
FT /tag= ac
FT /product= "ORF#21 protein"
FT /note= "encoded protein shown in AAW72187"
FT CDS 33140..34984
FT /tag= ad
FT /product= "ORF#22a protein"
FT /note= "encoded protein shown in AAW72226"
FT CDS 33386..34984
FT /tag= ae
FT /product= "ORF#22b protein"
FT /note= "encoded protein shown in AAW72227"
FT CDS complement (35205..37721)
FT /tag= af
FT /product= "ORF#23 protein"
FT /note= "encoded protein shown in AAW72188"
FT CDS complement (38058..39188)
FT /tag= ag
FT /product= "ORF#24 protein"
FT /note= "encoded protein shown in AAW72189"
FT CDS 39090..39935
FT /tag= ah
FT /product= "ORF#25 protein"
FT /note= "encoded protein shown in AAW72190"
FT CDS 40216..41973
FT /tag= ai
FT /product= "ORF#26 protein"
FT /note= "encoded protein shown in AAW72191"
FT CDS 42206..44178
FT /tag= aj
FT /transl_except= (pos: 44063..44064, aa: Lys)
FT /product= "ORF#27 protein"
FT /note= "encoded protein shown in AAW72192"
FT CDS complement (44853..47297)
FT /tag= ak
FT /product= "ORF#28 protein"
FT /note= "encoded protein shown in AAW72193"
FT CDS 47122..47338
FT /tag= al
FT /product= "ORF#29 protein"
FT /note= "encoded protein shown in AAW72194"
FT CDS complement (47305..49662)
FT /tag= am
FT /product= "ORF#30 protein"
FT /note= "encoded protein shown in AAW72195"

```

```

FT CDS complement (50035..51666)
FT /tag= an
FT /product= "ORF#31 protein"
FT /note= "encoded protein shown in AAW72196"
FT CDS complement (51701..53575)
FT /tag= ao
FT /product= "ORF#32 protein"
FT /note= "encoded protein shown in AAW72197"
FT CDS 54393..58115
FT /tag= ap
FT /product= "ORF#33 protein"
FT /note= "encoded protein shown in AAW72198"
FT CDS complement (58060..58977)
FT /tag= aq
FT /product= "ORF#34 protein"
FT /note= "encoded protein shown in AAW72199"
FT CDS complement (58970..60760)
FT /tag= ar
FT /product= "ORF#35 protein"
FT /note= "encoded protein shown in AAW72200"
FT CDS 60759..61151
FT /tag= as
FT /product= "ORF#36 protein"
FT /note= "encoded protein shown in AAW72201"
FT CDS 61241..62071
FT /tag= at

Query Match 56.0%; Score 14; DB 19; Length 117213;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 gcagcaggtgcac 16
DB 40479 gcagcaggtgcac 40492

RESULT 36
AAV19941/C
ID AAV19941 standard; DNA; 137507 BP.
AC AAV19941;
XX
XX 03-AUG-1998 (first entry)
DE KSHV long unique coding region and terminal repeat.
XX
XX KSHV; HHV8; human herpes virus 8; macrophage inflammatory protein II;
XX interleukin-6; IL-6; interferon regulatory factor; rheumatoid arthritis;
XX complement-binding protein; glycoprotein; capsid protein IV; infection;
XX immediate early protein; Kaposi's sarcoma; protective vaccine; lymphoma;
XX lymphoproliferative disease; leukaemia; splenomegaly; mycosis fungoides;
XX HIV immune status; anti-inflammatory agent; therapy; ds.
OS Kaposi's sarcoma-associated herpes virus.
XX
FH Key location/Qualifiers
FT 1142..2794
FT /tag= a
FT /product= complement-binding protein
FT 8699..11236
FT /tag= b
FT /product= glycoprotein B
FT complement (17261..17875)
FT /tag= c
FT /product= interleukin 6
FT complement (21548..21832)
FT /tag= d
FT /product= macrophage inflammatory protein II
FT complement (27137..27424)
FT /tag= e
FT /product= interferon regulatory factor 1
FT 2861..29741
FT /tag= f

```

```

FT      CDS      /product= protein T1.1
FT      CDS      complement (58976..60175)
FT      CDS      /tag= g
FT      CDS      /product= glycoprotein M
FT      CDS      complement (69412..69915)
FT      CDS      /tag= h
FT      CDS      /product= glycoprotein L
FT      CDS      complement (88410..88910)
FT      CDS      /tag= i
FT      CDS      /product= interferon regulatory factor 2
FT      CDS      89600..90541
FT      CDS      /tag= j
FT      CDS      /product= interferon regulatory factor 3
FT      CDS      90173..90643
FT      CDS      /tag= k
FT      CDS      /product= glycoprotein X
FT      CDS      complement (93636..94127)
FT      CDS      /tag= l
FT      CDS      /product= interferon regulatory factor 4
FT      CDS      complement (111931..112443)
FT      CDS      /tag= m
FT      CDS      /product= capsid protein IV
FT      CDS      complement (123808..127296)
FT      CDS      /tag= n
FT      CDS      /product= immediate early protein

```

```

XX      CDS      WO9804576-A1.
XX      CDS      05-FEB-1998.
XX      CDS      22-JUL-1997; 97WO-US13346.
XX      CDS      29-NOV-1996; 96US-0757669.
XX      CDS      25-JUL-1996; 96US-0686243.
XX      CDS      25-JUL-1996; 96US-0686349.
XX      CDS      25-JUL-1996; 96US-0686350.
XX      CDS      25-JUL-1996; 96US-0687253.
XX      CDS      25-JUL-1996; 96US-0688814.
XX      CDS      05-SEP-1996; 96US-0708678.
XX      CDS      10-OCT-1996; 96US-0728323.
XX      CDS      13-NOV-1996; 96US-0747887.
XX      CDS      13-NOV-1996; 96US-0748640.
XX      CDS      (UYCO ) UNIV COLUMBIA NEW YORK.
XX      CDS      Bohenzky RA, Chang Y, Edelman IS, Moore PS, Russo JU;
XX      CDS      WPI: 1998-130615/12.

```

XX CDS New nucleic acid encoding Kaposi's sarcoma associated herpes virus
 XX CDS proteins - useful for, e.g. detecting levels of HHV8 in, and
 XX CDS preparation of vaccines for treatment of, HIV patients

XX CDS Example 2: Page 135-203; 230pp; English.

XX CDS This sequence represents the long unique region and terminal repeat of
 CC the Kaposi's sarcoma-associated herpes virus (KSHV). KSHV is also known
 CC as human herpes virus 8 (HHV8). This sequence contains the DNAs of the
 CC invention which encode KSHV polypeptides selected from: (a) viral
 CC macrophage inflammatory protein (MIP) II; (b) viral interleukin-6 (IL-6);
 CC (c) viral IRF 1; (d) complement-binding protein; glycoproteins B, M or L;
 CC (e) capsid protein IV encoded by ORF65; and (e) immediate early protein
 CC encoded by ORF73. Labelled probes for the nucleic acid, proteins encoded
 CC by II, and antibodies (Ab) specific for the nucleic acid, proteins encoded
 CC by IIV, and specifically for the diagnosis of Kaposi's sarcoma, in body
 CC fluids or tissue samples. HHV8 infections can be treated with antisense
 CC or triplex forming molecules or agents that bind specifically to the
 CC protein. Ab may be used for prophylaxis or treatment of HHV8 infection,
 CC while the protein can be used in protective vaccines. Ab may also be used
 CC to differentiate between lymphomas, and HHV8 may be implicated in many
 CC other lymphoproliferative diseases such as lymphomas, leukaemia,
 CC splenomegaly and mycosis fungoides. Cells and animals containing the
 CC nucleic acid are useful for drug screening. HHV8-derived peptides can be

CC used as targets for antiviral drugs, e.g. dihydrofolate reductase gene
 CC can be inhibited with methotrexate. These can also be used to determine
 CC the immune status of a patient infected with HIV. HHV8 derived protein
 CC viral MIP II may be used as an anti-inflammatory agent for,
 CC e.g. treating rheumatoid arthritis. This sequence is stated as containing
 CC 81 open reading frames.

XX CDS 81 open reading frames.

XX CDS Sequence 137507 BP; 32579 A; 37795 C; 35758 G; 31375 T; 0 other;

XX CDS Query Match 56.0%; Score 14; DB 19; Length 137507;

XX CDS Best Local Similarity 100.0%; Pred. No. 72;

XX CDS Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

XX CDS 11 tgcacgcacagat 24

XX CDS Db 37158 TGCATCGCAGCAT 37145

XX CDS RESULT 37

XX CDS AAF21608/c

XX CDS ID AAF21608 standard; DNA; 349980 BP.

XX CDS AC AAF21608;

XX CDS DT 13-MAR-2001 (first entry)

XX CDS DE Neisseria meningitidis B nucleotide sequence SEQ ID NO:109.

XX CDS KW Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;

XX CDS KW diagnosis; antigen; detection; infection; gene therapy; antibacterial;

XX CDS KW ds.

XX CDS OS Neisseria meningitidis.

XX CDS PN WO200006791-A1.

XX CDS PD 09-NOV-2000.

XX CDS PE 08-MAR-2000; 2000WO-US05928.

XX CDS PR 30-APR-1999; 99US-0132068.

XX CDS PR 08-OCT-1999; 99WO-US23573.

XX CDS PR 28-FEB-2000; 2000CB-0004695.

XX CDS PA (CHIR) CHIRON CORP.

XX CDS PA (GENO-) INST GENOMIC RES.

XX CDS PI Pizza M, Hickey E, Peterson J, Tettelin H, Venter JC, Masignani V;

XX CDS PI Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V, Rappuoli R;

XX CDS PI Fraser CM, Grandi G.

XX CDS WPI: 2000-647603/62.

XX CDS Neisseria meningitidis B full length genome sequence and open reading

XX CDS frames are used to detect, treat and prevent Neisseria infections -

XX CDS Claim 7: Appendix A: 692pp; English.

XX CDS The present invention describes the full length genome of

XX CDS Neisseria meningitidis B (NMB). The sequences in AAF21544 and AAF21607

XX CDS CC to AAF21613 represent fragments of the NMB genomic sequence, as the

XX CDS sequence was too long to go in a record on its own it was split into 8

XX CDS CC sequences which overlap each other at the beginning and end of each

XX CDS CC sequence by 49980 bp (i.e. the last 49980 bp of AAF21544 is repeated at

XX CDS CC the beginning of AAF21607, the last 49980 bp of AAF21607 are repeated at

XX CDS CC the beginning of AAF21608, and so on). AAF21545 to AAF21588 encode the

XX CDS CC Neisseria proteins given in AAB58550 to AAB58593, and AAF21589 to

XX CDS CC AAF21606 represent PCR primers which are used in the exemplification of

XX CDS CC the present invention. The NMB genome and fragments from it have

XX CDS CC antibacterial activity, and can be used in vaccines and gene therapy.

XX CDS CC Neisseria nucleic acids, proteins and/or antibodies which binds to the

XX CDS CC proteins can be used in compositions for treating or preventing infection

CC due to Neisserial bacteria or as a diagnostic reagent for detecting the
CC presence of Neisserial bacteria or of antibodies raised to Neisserial
CC bacteria. Computers, computer memory, computer storage medium or computer
CC databases can be used in a search to identify open reading frames (ORFs)
CC or coding sequences within the NMB genome. The DNA sequences provide
CC further opportunities to find antigenic or immunogenic proteins which are
CC more effective in vaccines than the outer membrane proteins currently
CC used.

XX Sequence 349980 BP; 82523 A; 82940 C; 96712 G; 87805 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 349980;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gcgcagcaggtgca 14
|||||
Db 141640 GCGCAGCAGTCA 141627

RESULT 38

AAQ41769
ID AAQ41769 standard; DNA: 21 BP.

XX AC AAQ41769;

XX DT 03-SEP-1993 (first entry)

XX DE Myc EMSA probe production primer MLC-A.

XX KM Myc: c-myc; mammalian; E box; cancer; therapy; C1; C2; C2'; complex;

XX KM homo-oligomer; hetero-oligomer; myogenin; Max; oncoprotein; primer;

XX KM probe; electrophoretic mobility shift assay; EMSA; ss.

XX OS Synthetic.

XX FT Key Location/Qualifiers

XX FT protein_bind 13..18 /tag- a

XX PN MO9308701-A. /note- "E-Box core sequence"

XX PD 13-MAY-1993.

XX PF 09-OCT-1992; 92WO-0508603.

XX PR 30-OCT-1991; 91US-0785567.

XX PA (GEHO) GEN HOSPITAL CORP.

XX PI Kingston RE, Papoulas O;

XX DR WPI; 1993-167291/20.

XX PT Prod. of c-myc protein from mammalian cells - and detection of c-

XX PT Myc inhibitors for use in cancer therapy

XX PS Disclosure; Page 37; 101pp; English.

XX CC The sequences given in AAQ41764-75 are primers which were used to create

CC double stranded probes for use in an electrophoretic mobility shift

CC assay (EMSA) to isolate Myc binding sites from random sequences. The

CC isolated sequences contain the central E box core of CACGTG which

CC binds very weakly with Myc homo-oligomers (C1 complex), but more

CC tightly with Myc hetero-oligomers (C2 complex). The C2 complex

CC requires a 26-29 kd factor in addition to Myc. The additional factor

CC copurifies with Myc and resembles Max protein. A second copurifying

CC 40-50 kd factor has been identified (forming C2' complex). Sites

CC selected by the C2' complex contain the core CACGTG which bears

CC remarkable homology to a myogenin binding site (see AAQ41763).

CC Oligonucleotides containing the E box can be used in the purification

CC of Myc from a mammalian source. See also AAQ41761-861. The isolated

CC target sequences may be used in a method to inhibit c-myc oncoprotein

CC activity.

XX Sequence 21 BP; 4 A; 6 C; 8 G; 3 T; 0 other;

Query Match 52.0%; Score 13; DB 14; Length 21;
Best Local Similarity 100.0%; Pred. No. 2,8e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 gcgcagcaggtgca 14
|||||
Db 8 gcgcagcaggtgca 20

RESULT 39

AAAX3237
ID AAAX3237 standard; DNA: 21 BP.

XX AC AAAX3237;

XX DT 25-JUN-1999 (first entry)

XX DE Wheat viviparous 1 (tavp1) primer #4.

XX KM Wheat; oat; viviparous 1; VPI; afVPI; tavp1; maize; detection; PHS;

XX KM pre-harvest sprouting; dormant; germination; crop plant; primer; ss.

XX OS Synthetic.

XX FT Trifolium aestivum.

XX PN WO9915667-A1.

XX PD 01-APR-1999.

XX PF 18-SEP-1998; 98WO-GB02835.

XX PR 19-SEP-1997; 97GB-0020060.

XX PA (PLAN-) PLANT BIOSCIENCE LTD.

XX PI Flinham JE, Gale MD, Holdsworth MD;

XX DR WPI; 1999-244424/20.

XX PT New isolated oat and wheat VPI genes, used, e.g. to impose

XX PT sufficient dormancy to avoid pre-harvest sprouting

XX PS Claim 56; Page 89; 120pp; English.

XX CC The present sequence represents a primer for the wheat viviparous 1 (VPI)

CC gene, which keeps mature seeds dormant and inhibits germination. The

CC present invention describes genes which are homologues of the maize

CC Viviparous 1 gene, obtained from oat Avena fatua and wheat which encode

CC polypeptides designated afVPI and tavp1 respectively. The VPI activity

CC keeps mature seeds dormant and inhibits germination and can be used to

CC maintain or impose sufficient intensity and duration of dormancy to

CC avoid pre-harvest sprouting (PHS) before harvest. The products can be

CC used in the production of transformed crop plants having desirable

CC primary or secondary dormancy, or after-ripening properties, and in

CC particular may be resistant to PHS.

XX Sequence 21 BP; 4 A; 5 C; 9 G; 3 T; 0 other;

Query Match 52.0%; Score 13; DB 20; Length 21;
Best Local Similarity 100.0%; Pred. No. 2,8e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 gcgcagcaggtgcat 15
|||||
Db 4 gcgcagcaggtgcat 16

```

RESULT 40
AAV1150/c
ID AAV1150 standard; DNA; 22 BP.
XX
AC AAV1150;
XX
DT 14-JUL-1998 (first entry)
XX
DE Oligonucleotide #3 for nucleic acid assay.
XX
KW Nucleic acid assay; hybridising; probe; target; duplex; detection; ss.
XX
OS Synthetic.
XX
PN DE19633436-A1.
XX
PD 26-FEB-1998.
XX
PF 20-AUG-1996; 96DE-1033436.
XX
PR 20-AUG-1996; 96DE-1033436.
XX
PA (BOEF ) BOEHRINGER MANNHEIM GMBH.
XX
PI Bergmann F, Herrmann R, Kobold U;
XX
DR WPI; 1998-146248/14.
XX
PT Nucleic acid assay - based on mass of probe-target duplex
XX
PS Disclosure; Fig 1; 9pp; German.
XX
CC AAV1148-V1151 are oligonucleotides used in a novel nucleic acid assay.
CC This assay involves hybridising a probe to a target sequence of the
CC nucleic acid to form a duplex and degrading the nucleic acid to form
CC a fragment of a defined length containing the duplex, which is detected
CC on the basis of its mass. This method enables different nucleic acids to
CC be detected simultaneously does not require the use of labelled probes.
XX
SQ Sequence 22 BP; 5 A; 3 C; 7 G; 7 T; 0 other;

Query Match          52.0%; Score 13; DB 19; Length 22;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 catcgccagcatt 25
    |||||
Db 18 CATGCCGACGACTT 6

RESULT 41
AAA4502/c
ID AAA4502 standard; cDNA; 286 BP.
XX
AC AAA4502;
XX
DT 21-AUG-2000 (first entry)
XX
DE Human secreted expressed sequence tag SEQ ID NO:2077.
XX
KW Human; mouse; chicken; rat; secreted expressed sequence tag; SESTR;
KW expressed sequence tag; EST; probe; chemokine; proliferative;
KW immunomodulatory; haematopoietic; chemokine; analgesic; haemostatic;
KW thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal;
KW antiviral; antidiabetic; antiaslomatic; vulnerrary; antiparkinsonian;
KW antilicer; osteopathic; neuroprotective; nootropic; antiparatic;
KW cerebroprotective; anticonvulsant; antidepressant; gene therapy;
KW vaccine; autoimmune disorder; multiple sclerosis; allergic condition;
KW insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;
KW lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;

```

```

KW central nervous system disorder; Alzheimer's disease; stroke;
KW parkinson's disease; Huntington's disease; coagulation disorder;
KW haemophilia; thrombosis; inflammatory disorder; Crohn's disease;
KW tumour; infection; depression; psoriasis; ss.
XX
OS Homo sapiens.
XX
PN WO200021991-A1.
XX
PD 20-APR-2000.
XX
PF 15-OCT-1999; 99WO-US24206.
XX
PR 15-OCT-1998; 98US-010436.
XX
PA (GEM ) GENETICS INST INC.
XX
PI Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Bowman MR;
XX
DR WPI; 2000-317938/27.
XX
PT Isolated polynucleotides, and encoded proteins, comprising secreted
PT expressed sequence tags (SESTs), useful for treating various disorders
PT such as autoimmune, infectious, and central nervous system disorders -
XX
PS Claim 1; Page 723; 803pp; English.
XX
CC AAA43426 to AAA45925 represent specifically claimed secreted expressed
CC sequence tags (SESTs), isolated from human, mouse, chicken and rat
CC tissue sources. The SESTs can have a range of activities depending on
CC the tissues they were isolated from. The activities include:
CC chemotactic; proliferative; immunomodulatory; haematopoietic;
CC chemokine; analgesic; haemostatic; thrombolytic; antiinflammatory;
CC cytostatic; antibacterial; antifungal; antiviral; antidiabetic;
CC antiaslomatic; vulnerrary; antilicer; osteopathic; neuroprotective;
CC nootropic; antiparkinsonian; antiparatic; cerebroprotective;
CC anticonvulsant; and antidepressant. The SESTs can be used for gene
CC therapy and in vaccines. The SESTs are useful as probes for the
CC identification and isolation of full-length cDNAs and genomic DNA
CC molecules which correspond to the SESTs. Proteins encoded by the SESTs
CC are useful in assays for determining biological activity and raising
CC antibodies. They may be useful for treatment of autoimmune disorders
CC (multiple sclerosis, insulin dependent diabetes), allergic conditions
CC (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,
CC osteoporosis, osteoarthritis, central nervous system disorders
CC (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation
CC disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's
CC disease), tumours, bacterial, fungal or viral infections, depression and
CC psoriasis. AAA45926 to AAA45931 represent linker variants which are given
CC in the exemplification of the present invention.
XX
SQ Sequence 286 BP; 44 A; 84 C; 69 G; 89 T; 0 other;

Query Match          52.0%; Score 13; DB 21; Length 286;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 gcaggtgcacgc 18
    |||||
Db 93 GCAGGTGCATGCC 81

RESULT 42
AAZ14744/c
ID AAZ14744 standard; cDNA; 300 BP.
XX
AC AAZ14744;
XX
DT 12-OCT-1999 (first entry)
XX
DE Human gene expression product cDNA sequence SEQ ID NO:2213.

```

```

XX Human; gene: gene expression product; diagnosis: therapy; probe;
KW detection; mapping: tissue typing; profiling: forensic; cancer;
XX genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
XX Homo sapiens.
XX WO9338972-A2.
XX
XX 05-AUG-1999.
XX
XX 28-JAN-1999; 99WO-US01619.
XX
XX 03-APR-1998; 98US-0080666.
XX 28-JAN-1998; 98US-0072910.
XX 24-FEB-1998; 98US-0075954.
XX 31-MAR-1998; 98US-0080114.
XX 03-APR-1998; 98US-0080515.
XX
XX (CHIR ) CHIRON CORP.
XX (HYSE-) HYSEQ INC.
XX
XX Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;
XX Escobedo J, Kassam A, Kennedy GC, Kita D, Labat I;
XX Jones WL, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
XX Stache-Crain B, Sudduth-Klinger J, Williams LT;
XX WPI: 1999-494092/41.
XX
XX Novel human genes and their expression products which are
XX differentially expressed in different cell types
XX
XX Claim 1; Page 1159; 2479pp; English.
XX
XX The present invention describes a library of human polynucleotides
XX comprising the sequences given in AA212532 to AA217779. Also described is
XX a method of detecting differentially expressed genes correlated with the
XX cancerous state of a mammalian cell, comprising detecting at least one
XX differentially expressed gene product in a test sample from a cell
XX suspected of being cancerous, where the gene product is encoded by one
XX of the 5248 polynucleotide sequences given in AA212532 to AA217779. The
XX polynucleotides can be used as a source of primers and probes, which can
XX be used for a variety of purpose, e.g. detection of expression levels,
XX mapping, tissue typing or profiling, forensics, genetic analysis and
XX detection of polymorphisms. Polypeptides encoded by the polynucleotides
XX can be used for raising antibodies for experimental, diagnostic and
XX therapeutic purposes. The polynucleotides may also be used to construct
XX arrays for diagnostics (which may be used to determine function of an
XX encoded protein); and to detect differences in expression levels between
XX two cells (e.g. to identify abnormal or diseased tissue in a human, to
XX identify a genetic predisposition or susceptibility to a disease such as
XX cancer). The polynucleotides of the invention are especially used in the
XX diagnosis, prognosis and management of colorectal cancer, breast cancer,
XX and lung cancer. The polynucleotides can also be used to screen for
XX peptide analogues and antagonists.
XX
XX Sequence 300 BP; 83 A; 62 C; 70 G; 85 T; 0 other:
XX
XX Query Match 52.0%; Score 13; DB 20; Length 300;
XX Best Local Similarity 100.0%; Pred. No. 2.7e+02;
XX Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 12 gcacgcacgcat 24
XX ||||||||||||
XX Db 174 GCATGCCAGCAT 162
XX
XX RESULT 43
XX AA214726
XX ID AA214726 standard; CDNA: 300 BP.
XX

```

```

AC AA214726;
XX
XX 12-OCT-1999 (first entry)
XX
XX Human gene expression product cDNA sequence SEQ ID NO:2195.
XX
XX Human; gene: gene expression product; diagnosis: therapy; probe;
KW detection; mapping: tissue typing; profiling: forensic; cancer;
XX genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
XX Homo sapiens.
XX WO9338972-A2.
XX
XX 05-AUG-1999.
XX
XX 28-JAN-1999; 99WO-US01619.
XX
XX 03-APR-1998; 98US-0080666.
XX 28-JAN-1998; 98US-0072910.
XX 24-FEB-1998; 98US-0075954.
XX 31-MAR-1998; 98US-0080114.
XX 03-APR-1998; 98US-0080515.
XX
XX (CHIR ) CHIRON CORP.
XX (HYSE-) HYSEQ INC.
XX
XX Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;
XX Escobedo J, Kassam A, Kennedy GC, Kita D, Labat I;
XX Jones WL, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
XX Stache-Crain B, Sudduth-Klinger J, Williams LT;
XX WPI: 1999-494092/41.
XX
XX Novel human genes and their expression products which are
XX differentially expressed in different cell types
XX
XX Claim 1; Page 1155; 2479pp; English.
XX
XX The present invention describes a library of human polynucleotides
XX comprising the sequences given in AA212532 to AA217779. Also described is
XX a method of detecting differentially expressed genes correlated with the
XX cancerous state of a mammalian cell, comprising detecting at least one
XX differentially expressed gene product in a test sample from a cell
XX suspected of being cancerous, where the gene product is encoded by one
XX of the 5248 polynucleotide sequences given in AA212532 to AA217779. The
XX polynucleotides can be used as a source of primers and probes, which can
XX be used for a variety of purpose, e.g. detection of expression levels,
XX mapping, tissue typing or profiling, forensics, genetic analysis and
XX detection of polymorphisms. Polypeptides encoded by the polynucleotides
XX can be used for raising antibodies for experimental, diagnostic and
XX therapeutic purposes. The polynucleotides may also be used to construct
XX arrays for diagnostics (which may be used to determine function of an
XX encoded protein); and to detect differences in expression levels between
XX two cells (e.g. to identify abnormal or diseased tissue in a human, to
XX identify a genetic predisposition or susceptibility to a disease such as
XX cancer). The polynucleotides of the invention are especially used in the
XX diagnosis, prognosis and management of colorectal cancer, breast cancer,
XX and lung cancer. The polynucleotides can also be used to screen for
XX peptide analogues and antagonists.
XX
XX Sequence 300 BP; 80 A; 64 C; 63 G; 93 T; 0 other:
XX
XX Query Match 52.0%; Score 13; DB 20; Length 300;
XX Best Local Similarity 100.0%; Pred. No. 2.7e+02;
XX Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 6 gcagtgcatgcgc 18
XX ||||||||||||
XX Db 180 gcagtgcatgcgc 192
XX

```

```

RESULT 44
AAZ12980/C
ID AAZ12980 standard; cDNA: 300 BP.
AC AAZ12980;
XX
XX 12-OCT-1999 (first entry)
DE Human gene expression product cDNA sequence SEQ ID NO:449.
XX
XX Human; gene; gene expression product; diagnosis; therapy; probe;
KW detection; mapping; tissue typing; profiling; forensic; cancer;
KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
XX
OS Homo sapiens.
XX
XX WO938972-A2.
XX
XX 05-AUG-1999.
XX
XX 28-JAN-1999; 99WO-US01619.
XX
XX 03-APR-1998; 98US-0080666.
XX
XX 28-JAN-1998; 98US-0072910.
XX
XX 24-FEB-1998; 98US-0075954.
XX
XX 31-MAR-1998; 98US-0080114.
XX
XX 03-APR-1998; 98US-0080515.
XX
XX (CHIR ) CHIRON CORP.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;
PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;
PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;
PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
PI Stache-Crain B, Sudduth-Klinger J, Williams LT;
XX
XX WPI: 1999-494092/41.
XX
XX Novel human genes and their expression products which are
XX differentially expressed in different cell types
XX
XX Claim 1; Page 764; 2479pp; English.
XX
XX The present invention describes a library of human polynucleotides
XX comprising the sequences given in AAZ12532 to AAZ17779. Also described is
XX a method of detecting differentially expressed genes correlated with the
XX cancerous state of a mammalian cell, comprising detecting at least one
XX differentially expressed gene product in a test sample from a cell
XX suspected of being cancerous, where the gene product is encoded by one
XX of the 5248 polynucleotide sequences given in AAZ12532 to AAZ17779. The
XX polynucleotides can be used as a source of primers and probes, which can
XX be used for a variety of purpose, e.g. detection of expression levels,
XX mapping, tissue typing or profiling, forensics, genetic analysis and
XX detection of polymorphisms. Polypeptides encoded by the polynucleotides
XX can be used for raising antibodies for experimental, diagnostic and
XX therapeutic purposes. The polynucleotides may also be used to construct
XX arrays for diagnostics (which may be used to determine function of an
XX encoded protein); and to detect differences in expression levels between
XX two cells (e.g. to identify abnormal or diseased tissue in a human, to
XX identify a genetic predisposition or susceptibility to a disease such as
XX cancer). The polynucleotides of the invention are especially used in the
XX diagnosis, prognosis and management of colorectal cancer, breast cancer,
XX and lung cancer. The polynucleotides can also be used to screen for
XX peptide analogues and antagonists.
XX
XX Sequence 300 BP; 69 A; 89 C; 77 G; 65 T; 0 other;

```

```

Query Match 52.0%; Score 13; DB 20; Length 300;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 2 ccgacgcaggtgca 14
Db 32 CGCAGCAGGTGCA 20

```

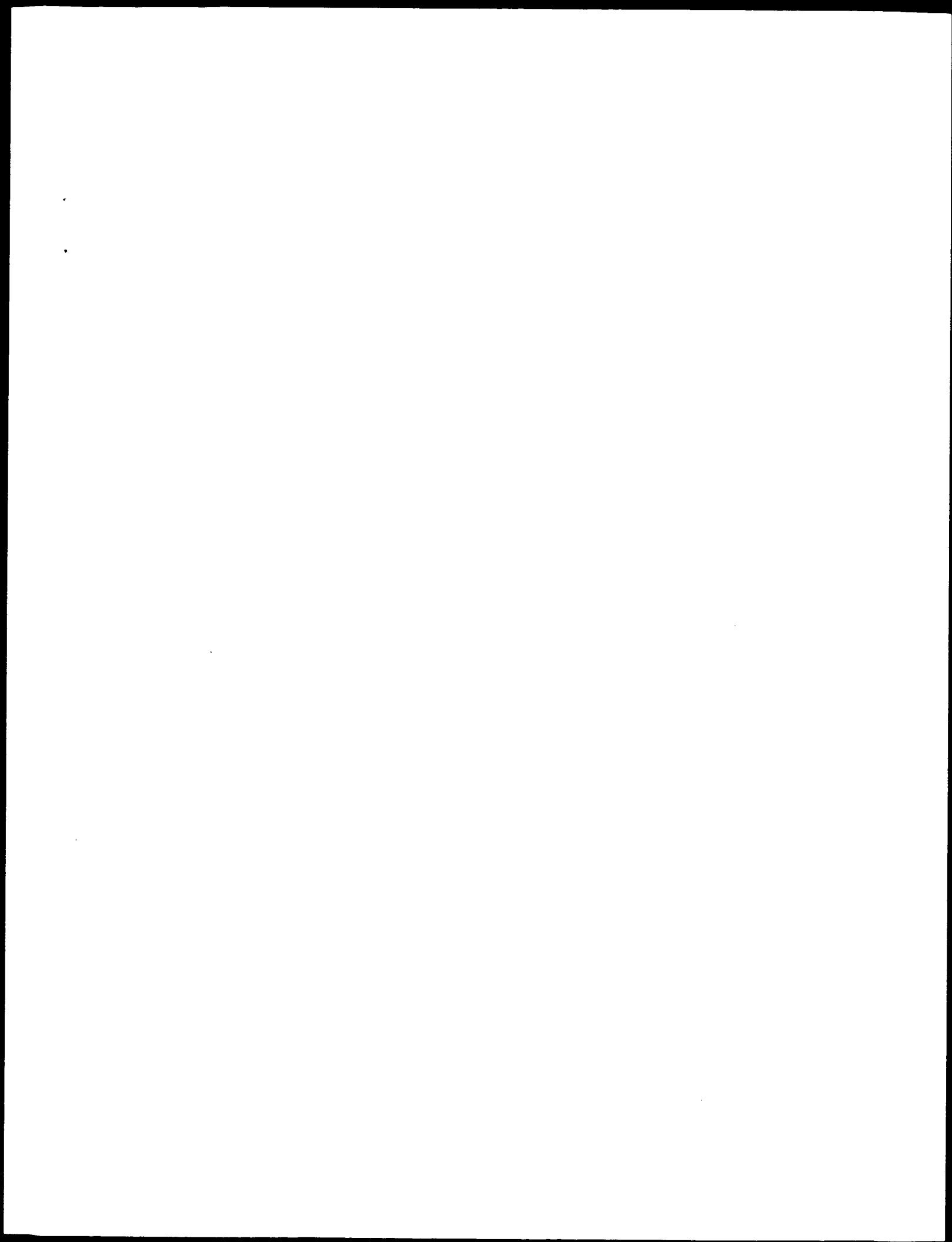
```

RESULT 45
AAZ12981/C
ID AAZ12981 standard; cDNA: 300 BP.
AC AAZ12981;
XX
XX 12-OCT-1999 (first entry)
DE Human gene expression product cDNA sequence SEQ ID NO:450.
XX
XX Human; gene; gene expression product; diagnosis; therapy; probe;
KW detection; mapping; tissue typing; profiling; forensic; cancer;
KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
XX
XX Homo sapiens.
XX
XX WO938972-A2.
XX
XX 05-AUG-1999.
XX
XX 28-JAN-1999; 99WO-US01619.
XX
XX 03-APR-1998; 98US-0080666.
XX
XX 28-JAN-1998; 98US-0072910.
XX
XX 24-FEB-1998; 98US-0075954.
XX
XX 31-MAR-1998; 98US-0080114.
XX
XX 03-APR-1998; 98US-0080515.
XX
XX (CHIR ) CHIRON CORP.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;
PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;
PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;
PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
PI Stache-Crain B, Sudduth-Klinger J, Williams LT;
XX
XX WPI: 1999-494092/41.
XX
XX Novel human genes and their expression products which are
XX differentially expressed in different cell types
XX
XX Claim 1; Page 764; 2479pp; English.
XX
XX The present invention describes a library of human polynucleotides
XX comprising the sequences given in AAZ12532 to AAZ17779. Also described is
XX a method of detecting differentially expressed genes correlated with the
XX cancerous state of a mammalian cell, comprising detecting at least one
XX differentially expressed gene product in a test sample from a cell
XX suspected of being cancerous, where the gene product is encoded by one
XX of the 5248 polynucleotide sequences given in AAZ12532 to AAZ17779. The
XX polynucleotides can be used as a source of primers and probes, which can
XX be used for a variety of purpose, e.g. detection of expression levels,
XX mapping, tissue typing or profiling, forensics, genetic analysis and
XX detection of polymorphisms. Polypeptides encoded by the polynucleotides
XX can be used for raising antibodies for experimental, diagnostic and
XX therapeutic purposes. The polynucleotides may also be used to construct
XX arrays for diagnostics (which may be used to determine function of an
XX encoded protein); and to detect differences in expression levels between
XX two cells (e.g. to identify abnormal or diseased tissue in a human, to
XX identify a genetic predisposition or susceptibility to a disease such as
XX cancer). The polynucleotides of the invention are especially used in the
XX diagnosis, prognosis and management of colorectal cancer, breast cancer,
XX and lung cancer. The polynucleotides can also be used to screen for
XX peptide analogues and antagonists.
XX
XX Sequence 300 BP; 69 A; 88 C; 77 G; 66 T; 0 other;

```

Query Match 52.0%; Score 13; DB 20; Length 300;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 cgcagcaggtgca 14
|||||
Db 32 CGCAGCAGGTGCA 20

Search completed: October 9, 2001, 15:52:25
Job time: 15106 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 15:55:27 ; Search time 218.82 seconds
(without alignments)
21.629 Million cell updates/sec

Title: US-09-396-196f-5

Perfect score: 25

Sequence: 1 ggcgcagcagtgatcgcgcagcatt 25

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 324599 seqs, 9465562 residues

Word size : 9

Total number of hits satisfying chosen parameters: 11705

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

Issued Patents NA: *
1: /cgn2_6/prodata/2/ina/5A_COMB.seq: *
2: /cgn2_6/prodata/2/ina/5B_COMB.seq: *
3: /cgn2_6/prodata/2/ina/6A_COMB.seq: *
4: /cgn2_6/prodata/2/ina/5B_COMB.seq: *
5: /cgn2_6/prodata/2/ina/PCRTUS_COMB.seq: *
6: /cgn2_6/prodata/2/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	100.0	1041	2	US-08-401-068-7
2	25	100.0	1041	2	US-08-846-338-7
3	25	100.0	5872	3	US-08-411-768B-1
4	25	100.0	5872	3	US-08-411-768B-6
5	14	56.0	718	4	US-09-142-078-45
6	14	56.0	2193	4	US-08-420-235B-16
7	14	56.0	2193	3	US-08-793-624-16
8	14	56.0	2193	5	PCRT-US95-10194-16
9	14	56.0	2973	2	US-08-343-101A-5
10	14	56.0	2973	3	US-09-183-688-5
11	14	56.0	4182	1	US-07-973-257-1
12	14	56.0	5362	2	US-08-853-310-3
13	14	56.0	20710	1	US-08-420-235B-1
14	14	56.0	20710	5	PCRT-US95-10194-1
15	14	56.0	20710	5	PCRT-US95-10194-1
16	14	56.0	35100	2	US-08-770-379-18
17	14	56.0	35100	2	US-08-757-669A-18
18	14	56.0	304	4	US-09-060-756-726
19	13	52.0	311	4	US-09-060-756-726
20	13	52.0	411	2	US-08-640-978C-2
21	13	52.0	417	2	US-08-640-978C-4
22	13	52.0	476	4	US-08-836-075A-51
23	13	52.0	476	4	US-08-998-416-978
24	13	52.0	1128	1	US-07-872-678A-11
25	13	52.0	1202	2	US-08-428-414A-1
26	13	52.0	1365	4	US-09-319-892-1
27	13	52.0	1624	2	US-07-862-588B-6

28	13	52.0	1775	2	US-07-862-588B-5	Sequence 5, Appl1
29	13	52.0	1875	3	US-09-258-373-21	Sequence 21, Appl1
30	13	52.0	1981	3	US-09-017-706-3	Sequence 3, Appl1
31	13	52.0	1981	3	US-09-017-706-4	Sequence 4, Appl1
32	13	52.0	1981	3	US-09-017-706-5	Sequence 5, Appl1
33	13	52.0	1981	3	US-09-017-706-6	Sequence 6, Appl1
34	13	52.0	1981	3	US-09-017-706-7	Sequence 7, Appl1
35	13	52.0	1981	3	US-09-017-706-8	Sequence 8, Appl1
36	13	52.0	2223	1	US-08-317-522A-8	Sequence 8, Appl1
37	13	52.0	2223	2	US-08-439-818A-8	Sequence 8, Appl1
38	13	52.0	2223	2	US-08-751-965-8	Sequence 8, Appl1
39	13	52.0	2223	2	US-08-738-975-8	Sequence 8, Appl1
40	13	52.0	2223	2	US-08-728-626-8	Sequence 8, Appl1
41	13	52.0	2223	3	US-08-808-599A-8	Sequence 8, Appl1
42	13	52.0	2290	1	US-08-584-226-1	Sequence 1, Appl1
43	13	52.0	2763	3	US-08-941-445A-16	Sequence 16, Appl1
44	13	52.0	3618	1	US-07-872-678A-36	Sequence 36, Appl1
45	13	52.0	3670	1	US-08-386-727-3	Sequence 3, Appl1

ALIGNMENTS

RESULT 1
US-08-401-068-7
; Sequence 7, Application US/08401068
; Patent No. 5859335
; GENERAL INFORMATION:
; APPLICANT: Patton, David
; TITLE OF INVENTION: Enhanced Biotin Biosynthesis in Plant Tissue
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/401,068
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/351,970
; FILING DATE: 08-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1041 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1038
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /product= "Biotin synthase"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
US-08-401-068-7

Query Match 100.0%; Score 25; DB 2; Length 1041;
Best Local Similarity 100.0%; Pred. No. 5.3e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcgcagcaggtgcgcgcgcatt 25
|||||
DB 79 GCGCAGCAGGTGCATCGCCAGCAT 103

RESULT 2

US-08-846-338-7
Sequence 7, Application US/08846338
Patent No. 5869719

GENERAL INFORMATION:

APPLICANT: Patton, David
TITLE OF INVENTION: Transgenic Plants Having Increased Biotin Content
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESS: No. 5869719artis Corporation
STREET: 520 White Plains Road, P.O. Box 2005
CITY: Tarrytown
STATE: NY
COUNTRY: USA
ZIP: 10591

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/846,338
FILING DATE:

CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SRO ID NO: 7;

SEQUENCE CHARACTERISTICS:

LENGTH: 1041 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
FEATURE:

NAME/KEY: CDS

LOCATION: 1..1038
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /product= "biotin synthase"
OTHER INFORMATION: /evidence= EXPERIMENTAL

US-08-846-338-7

Query Match 100.0%; Score 25; DB 2; Length 1041;
Best Local Similarity 100.0%; Pred. No. 5.3e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcgcagcaggtgcgcgcgcatt 25
|||||
DB 79 GCGCAGCAGGTGCATCGCCAGCAT 103

RESULT 3

US-08-411-768B-1
Sequence 1, Application US/08411768B
Patent No. 6083712

GENERAL INFORMATION:

APPLICANT: Olwen Birch
APPLICANT: Johann Brass
APPLICANT: Martin Fuhrmann

APPLICANT: Nicholas Shaw
TITLE OF INVENTION: Biotechnological Method
of Producing Biotin
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:

ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10112

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Wordperfect
SOFTWARE: Version 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/411,768B
FILING DATE: 31-March-95
CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: CH 3124/92
FILING DATE: 02-OCT-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: CH 2134/93
FILING DATE: 15-JUL-1993

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:
LENGTH: 5872 base pairs
TYPE: nucleic acid
STRANDEDNESS: double

TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: Escherichia coli
STRAIN: DSM498

IMMEDIATE SOURCE:

CLONE: pBO30A-15/9
FEATURE:

NAME/KEY: CDS

LOCATION: 117..1157
IDENTIFICATION METHOD: experimental

OTHER INFORMATION: /codon_start= 117
OTHER INFORMATION: /product= "Biotin synthase"

OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "biob"

OTHER INFORMATION: /number= 1
FEATURE:

NAME/KEY: CDS
LOCATION: 2295..3050

OTHER INFORMATION: /codon_start= 2295
OTHER INFORMATION: /function= "involved in pimeloyl-CoA synthesis"

OTHER INFORMATION: /product= "protein"
OTHER INFORMATION: /gene= "bioc"

OTHER INFORMATION: /number= 3
FEATURE:

NAME/KEY: CDS
LOCATION: 3750..5039

IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start= 3750
OTHER INFORMATION: /EC_number= 2.6.1.62

OTHER INFORMATION: /product= "DAPA synthase"
OTHER INFORMATION: /evidence= EXPERIMENTAL

OTHER INFORMATION: /gene= "bioA"
OTHER INFORMATION: /number= 5
OTHER INFORMATION: /standard_name=

OTHER INFORMATION: "S-Adenosyl-L-methionine:8-amino-7-oxononanoate
amino-transf."

FEATURE:

NAME/KEY: CDS

```

? LOCATION: 5098..5574
? IDENTIFICATION METHOD: experimental
? OTHER INFORMATION: /codon_start=5098
? OTHER INFORMATION: /function="unknown, involved in biotin synthesis"
? OTHER INFORMATION: /product="protein"
? OTHER INFORMATION: /evidence=EXPERIMENTAL
? OTHER INFORMATION: /gene="ORFI"
? FEATURE:
? NAME/KEY: -10_signal
? LOCATION: 45..49
? IDENTIFICATION METHOD: experimental
? OTHER INFORMATION: /evidence=EXPERIMENTAL
? OTHER INFORMATION: /standard_name="promoter ptac"
? FEATURE:
? NAME/KEY: -35_signal
? LOCATION: 23..28
? OTHER INFORMATION: /standard_name="promoter ptac"
? FEATURE:
? NAME/KEY: RBS
? LOCATION: 105..119
? IDENTIFICATION METHOD: experimental
? OTHER INFORMATION: /evidence=EXPERIMENTAL
? OTHER INFORMATION: /standard_name="biob RBS no.9"
? FEATURE:
? NAME/KEY: RBS
? LOCATION: 2284..2297
? OTHER INFORMATION: /standard_name="bioc RBS"
? FEATURE:
? NAME/KEY: RBS
? LOCATION: 3742..3752
? OTHER INFORMATION: /standard_name="bioc RBS"
? FEATURE:
? NAME/KEY: RBS
? LOCATION: 5088..5100
? OTHER INFORMATION: /standard_name="ORFI RBS"
? FEATURE:
? NAME/KEY: terminator
? LOCATION: 5583..5644
? OTHER INFORMATION: /standard_name="rho-independent
? OTHER INFORMATION: transcriptional terminator"
? FEATURE:
? NAME/KEY: stem_loop
? LOCATION: 5583..5605
? FEATURE:
? NAME/KEY: promoter
? LOCATION: 1..96
? IDENTIFICATION METHOD: experimental
? OTHER INFORMATION: /function="promoter ptac"
? OTHER INFORMATION: /evidence=EXPERIMENTAL
? PUBLICATION INFORMATION:
? DOCUMENT NUMBER: WO 87/01391 B1
? FILING DATE: 26-AUG-1986
? PUBLICATION DATE: 07-APR-1993
? US-08-411-768B-1

Query Match 100.0%; Score 25; DB 3; Length 5872;
Best Local Similarity 100.0%; Pred. No. 5.2e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ggcgcagcagtgatccgcacatt 25
    |||||||||||||||||||
Db 195 cgcgcagcagtgatccgcacatt 219

RESULT 4
US-08-411-768B-6
: Sequence 6, Application US/08411768B
: Patent No. 6083712
: GENERAL INFORMATION:
: APPLICANT: Olwen Birch
: APPLICANT: Johann Brass
:

? APPLICANT: Martin Fuhrmann
? APPLICANT: Nicholas Shaw
? TITLE OF INVENTION: Biotechnological Method
? TITLE OF INVENTION: of Producing Biotin
? NUMBER OF SEQUENCES: 19
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
? STREET: 30 Rockefeller Plaza
? CITY: New York
? STATE: New York
? COUNTRY: USA
? ZIP: 10112
? COMPUTER READABLE FORM:
? MEDIUM TYPE: floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Wordperfect
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/411,768B
? FILING DATE: 31-March-95
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: CH 3124/92
? FILING DATE: 02-OCT-1992
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: CH 2134/93
? FILING DATE: 15-JUL-1993
? INFORMATION FOR SEQ ID NO: 6:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 5872 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: double
? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
? HYPOTHETICAL: NO
? ORIGINAL SOURCE:
? ORGANISM: Escherichia coli
? STRAIN: DSM498
? IMMEDIATE SOURCE:
? CLONE: PHO30A15-9
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 1154..2308
? IDENTIFICATION METHOD: experimental
? OTHER INFORMATION: /codon_start=1154
? OTHER INFORMATION: /EC_number=2.3.1.47
? OTHER INFORMATION: /product="KAPA synthase"
? OTHER INFORMATION: /evidence=EXPERIMENTAL
? OTHER INFORMATION: /gene="biop"
? OTHER INFORMATION: /number=2
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 3043..3753
? IDENTIFICATION METHOD: experimental
? OTHER INFORMATION: /codon_start=3043
? OTHER INFORMATION: /EC_number=6.3.3.3
? OTHER INFORMATION: /product="PTB synthase"
? OTHER INFORMATION: /evidence=EXPERIMENTAL
? OTHER INFORMATION: /gene="biop"
? OTHER INFORMATION: /number=4
? FEATURE:
? NAME/KEY: RBS
? LOCATION: 1141..1156
? OTHER INFORMATION: /standard_name="biop RBS"
? FEATURE:
? NAME/KEY: RBS
? LOCATION: 3030..3045
? OTHER INFORMATION: /standard_name="biop RBS"
? PUBLICATION INFORMATION:
? DOCUMENT NUMBER: WO 87/01391 B1
```

Wed Oct 10 07:45:55 2001

FILING DATE: 26-AUG-1986
 PUBLICATION DATE: 07-APR-1993
 US-08-411-768B-6

Query Match 100.0%; Score 25; DB 3; Length 5872;
 Best Local Similarity 100.0%; Pred. No. 5.2e-05;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 gccgacagtgatccagcatt 25
 ||||||||||||||||||
 Db 195 gccgacagtgatccagcatt 219

RESULT 5
 US-09-142-078-45/c
 Sequence 45, Application US/09142078
 Patent No. 6172041

GENERAL INFORMATION:
 APPLICANT: McCabe, R. Tyler
 APPLICANT: Zhou, Li-Ming
 APPLICANT: Layer, Richard T.
 TITLE OF INVENTION: Use of Conantokins
 NUMBER OF SEQUENCES: 71
 CORRESPONDENCE ADDRESS:

ADDRESSEE: Rothwell, Figg, Ernst & Kurtz, P.C.
 STREET: 555 Thirteenth Street, N.W., Suite 701-E
 CITY: Washington
 STATE: D.C.
 COUNTRY: USA
 ZIP: 20004

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/142,078
 FILING DATE: 10-FEB-1993

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: WO US97/12652
 FILING DATE: 21-JUL-1997
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/672,377
 FILING DATE: 06-DEC-1996

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/684,750
 FILING DATE: 22-JUL-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Ihnen, Jeffrey L.
 REGISTRATION NUMBER: 28,957
 REFERENCE/DOCKET NUMBER: 2314-135.A

TELECOMMUNICATION INFORMATION:
 TELEPHONE: 202-783-6040
 TELEFAX: 202-783-6031
 INFORMATION FOR SEQ ID NO: 45:

SEQUENCE CHARACTERISTICS:
 LENGTH: 718 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (cDNA)
 ORIGINAL SOURCE:
 ORGANISM: Cornus geographus
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 110..409
 US-09-142-078-45

Query Match 56.0%; Score 14; DB 4; Length 718;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 7 caggtgcatcgcca 20
 ||||||||||||
 Db 118 CAGGTGTCATCGCCA 105

RESULT 6
 US-08-420-235B-16/c
 Sequence 16, Application US/08420235B
 Patent No. 5801042

GENERAL INFORMATION:
 APPLICANT: Chang, Yuan
 APPLICANT: Moore, Patrick S.
 TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
 TITLE OF INVENTION: SEQUENCES AND USES THEREOF
 NUMBER OF SEQUENCES: 47
 CORRESPONDENCE ADDRESS:

ADDRESSEE: Cooper & Dunham LLP
 STREET: 1185 Avenue of the Americas
 CITY: New York
 STATE: New York
 COUNTRY: U.S.A.
 ZIP: 10036

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/420,235B
 FILING DATE:

CLASSIFICATION: 424
 ATTORNEY/AGENT INFORMATION:
 NAME: White, John P.
 REGISTRATION NUMBER: 28,678
 REFERENCE/DOCKET NUMBER: 45185-B
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 278-0400
 TELEFAX: (212) 391-0525

INFORMATION FOR SEQ ID NO: 16:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2193 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single

TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 HYPOTHETICAL: N
 ANTI-SENSE: N
 FEATURE:

NAME/KEY: CDS
 LOCATION: 1..2193
 OTHER INFORMATION:
 US-08-420-235B-16

Query Match 56.0%; Score 14; DB 1; Length 2193;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 11 tgcacgcagcatt 24
 ||||||||||||
 Db 46 TGCATGCCAGCAT 33

RESULT 7
 US-08-793-624-16/c
 Sequence 16, Application US/08793624C
 Patent No. 6150093

GENERAL INFORMATION:
 APPLICANT: Chang, Yuan
 APPLICANT: Moore, Patrick S.
 TITLE OF INVENTION: Unique Associated Kaposi's Sarcoma Virus Sequences And
 TITLE OF INVENTION: Uses Thereof

FILE REFERENCE: 45185-C-PCT-US/JPW
CURRENT APPLICATION NUMBER: US/08/793,624C
CURRENT FILING DATE: 1997-02-18
NUMBER OF SEQ ID NOS: 58
SOFTWARE: Patentln Ver. 2.1
SEQ ID NO 16
LENGTH: 2193
TYPE: DNA
ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-08-793-624-16

Query Match 56.0%; Score 14; DB 3; Length 2193;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 tgcacgccagcat 24
|||||
DB 46 TGCATGCCAGCAT 33

RESULT 8
PCT-US95-10194-16/c

Sequence 16, Application PC/TUS9510194

GENERAL INFORMATION:

APPLICANT: The Trustees of Columbia University in the City of New York

APPLICANT: City

TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS

TITLE OF INVENTION: SEQUENCES AND USES THEREOF

NUMBER OF SEQUENCES: 45

CORRESPONDENCE ADDRESS:

ADDRESSEE: Cooper & Dunham LLP

STREET: 1185 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10036

COMPUTER READABLE FORM:

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentln Release #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/10194

FILING DATE:

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: White, John P.

REGISTRATION NUMBER: 28,678

REFERENCE/DOCKET NUMBER: 45185-C-PCT/JPW/MS

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 278-0400

TELEFAX: (212) 391-0525

INFORMATION FOR SEQ ID NO: 16:

SEQUENCE CHARACTERISTICS:

LENGTH: 2193 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: N

ANTI-SENSE: N

FEATURE:

NAME/KEY: CDS

LOCATION: 1..2193

OTHER INFORMATION:

PCT-US95-10194-16

Query Match 56.0%; Score 14; DB 5; Length 2193;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 tgcacgccagcat 24
|||||
DB 46 TGCATGCCAGCAT 33

RESULT 9
US-08-343-101A-5/c

Sequence 5, Application US/08343101A

Patent No. 5830759

GENERAL INFORMATION:

APPLICANT: Chang, Yuan

APPLICANT: Moore, Patrick S.

TITLE OF INVENTION: Unique Associated Kaposi's Sarcoma

TITLE OF INVENTION: Virus Sequences And Uses Thereof

NUMBER OF SEQUENCES: 22

CORRESPONDENCE ADDRESS:

ADDRESSEE: Cooper & Dunham

STREET: 1185 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentln Release #1.24

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/343,101A

FILING DATE:

CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: White Esq., John P.

REGISTRATION NUMBER: 28,678

REFERENCE/DOCKET NUMBER: 45185-A

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-278-0400

TELEFAX: 212-391-0526

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 2973 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: N

ANTI-SENSE: N

US-08-343-101A-5

Query Match 56.0%; Score 14; DB 2; Length 2973;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 tgcacgccagcat 24
|||||
DB 2182 TGCATGCCAGCAT 2169

RESULT 10

US-09-183-688-5/c

Sequence 5, Application US/09183688

Patent No. 6093550

GENERAL INFORMATION:

APPLICANT: Chang, Yuan

APPLICANT: Moore, Patrick S.

TITLE OF INVENTION: Unique Associated Kaposi's Sarcoma

TITLE OF INVENTION: Virus Sequences And Uses Thereof

NUMBER OF SEQUENCES: 22

CORRESPONDENCE ADDRESS:

ADDRESSEE: Cooper & Dunham

STREET: 1185 Avenue of the Americas

CITY: New York

```
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.24
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/09/183,688
  FILING DATE:
  CLASSIFICATION:
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: 08/343,101
  FILING DATE:
  ATTORNEY/AGENT INFORMATION:
    NAME: White Esq., John P.
    REGISTRATION NUMBER: 28,678
    REFERENCE/DOCKET NUMBER: 45185-A
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: 212-278-0400
    TELEFAX: 212-391-0526
  INFORMATION FOR SEQ ID NO: 5:
    SEQUENCE CHARACTERISTICS:
      LENGTH: 2973 base pairs
      TYPE: nucleic acid
      STRANDEDNESS: single
      TOPOLOGY: linear
    MOLECULE TYPE: DNA (genomic)
    HYPOTHETICAL: N
    ANTI-SENSE: N
  US-09-183-688-5

Query Match
Best Local Similarity 56.0%; Score 14; DB 3; Length 2973;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 tgcacgcacgat 24
      |||||||
Db 2182 tgcacgcacgat 2169

RESULT 11
US-07-973-257-1/c
Sequence 1, Application US/07973257
Patent No. 5378620
GENERAL INFORMATION:
  APPLICANT: Keeler, Jr. Calvin L.
  APPLICANT: Dohms, John E.
  TITLE OF INVENTION: Gene Encoding Cytoadhesin
  TITLE OF INVENTION: Protein of Mycoplasma Gallisepticum and Its Use
  NUMBER OF SEQUENCES: 1
  CORRESPONDENCE ADDRESSES:
    ADDRESSEE: Connolly and Hutz
    STREET: 1220 Market Street
    CITY: Wilmington
    STATE: Delaware
    COUNTRY: U.S.A.
  ZIP: 19899
  COMPUTER READABLE FORM:
    MEDIUM TYPE: 3.5 inch diskette
    COMPUTER: IBM/PC or Compatible
    OPERATING SYSTEM: MS-DOS
    SOFTWARE: Wordperfect 5.1
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/07/973,257
    FILING DATE: 19921109
  CLASSIFICATION: 435
  PRIOR APPLICATION DATA: No. 5378820e
  INFORMATION FOR SEQ ID NO: 1:
    SEQUENCE CHARACTERISTICS:
      LENGTH: 4182 base pairs
```

```
TYPE: NUCLEIC ACID
STRANDEDNESS: double stranded
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Mycoplasma Gallisepticum
STRAIN: S6
US-07-973-257-1

Query Match
Best Local Similarity 56.0%; Score 14; DB 1; Length 4182;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 gcacgagtgac 16
      |||||||
Db 1260 GCACGAGTGACATC 1247

RESULT 12
US-08-853-310-3/c
Sequence 3, Application US/08853310
Patent No. 5948640
GENERAL INFORMATION:
  APPLICANT: Randazzo, Filippo
  TITLE OF INVENTION: Mammalian Additional Sex Combs (Asx) Acts as a Tumor Suppre
  NUMBER OF SEQUENCES: 4
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: Chiron Corporation
    STREET: 4560 Horton Street
    CITY: Emeryville
    STATE: California
    COUNTRY: U.S.A.
  ZIP: 94608
  COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: Patent Release #1.0, Version #1.30
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/853,310
    FILING DATE:
  CLASSIFICATION: 514
  ATTORNEY/AGENT INFORMATION:
    NAME: Gutth, Joseph H.
    REGISTRATION NUMBER: 31,261
    REFERENCE/DOCKET NUMBER: 1228,003
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: (510) 923-3888
    TELEFAX: (510) 655-3542
  INFORMATION FOR SEQ ID NO: 3:
    SEQUENCE CHARACTERISTICS:
      LENGTH: 5362 base pairs
      TYPE: nucleic acid
      STRANDEDNESS: single
      TOPOLOGY: linear
    MOLECULE TYPE: DNA (genomic)
  US-08-853-310-3

Query Match
Best Local Similarity 56.0%; Score 14; DB 2; Length 5362;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 gcacgagtgac 16
      |||||||
Db 3354 GCACGAGTGACATC 3341

RESULT 13
US-08-420-235B-1/c
```

```
; Sequence 1, Application US/08420235B
; Patent No. 5801042
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/420,235B
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 45185-B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20710 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: N
; ANTI-SENSE: N
US-08-420-235B-1
```

```
Query Match          56.0%; Score 14; DB 1; Length 20710;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 tgcacgcacgcat 24
      |||
Db 2138 TGCATGCCGACAT 2125

RESULT 14
US-08-793-624-1/c
; Sequence 1, Application US/08793624C
; Patent No. 6150093
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: Unique Associated Kaposi's Sarcoma Virus Sequences And
; TITLE OF INVENTION: Uses Thereof
; FILE REFERENCE: 45185-C-PCT-US/JPM
; CURRENT APPLICATION NUMBER: US/08/793,624C
; CURRENT FILING DATE: 1997-02-18
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1
; LENGTH: 20710
; TYPE: DNA
; ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-08-793-624-1
```

```
Query Match          56.0%; Score 14; DB 3; Length 20710;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 tgcacgcacgcat 24
      |||
Db 2138 TGCATGCCGACAT 2125
```

```
RESULT 15
PCT-US95-10194-1/c
; Sequence 1, Application PC/TUS9510194
; GENERAL INFORMATION:
; APPLICANT: The Trustees of Columbia University in the City of New York
; APPLICANT: City
; TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/10194
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 45185-C-PCT/JPM/MS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20710 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: N
; ANTI-SENSE: N
PCT-US95-10194-1
```

```
Query Match          56.0%; Score 14; DB 5; Length 20710;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 tgcacgcacgcat 24
      |||
Db 2138 TGCATGCCGACAT 2125

RESULT 16
US-08-770-379-18/c
; Sequence 18, Application US/08770379
; Patent No. 5849564
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A.
; APPLICANT: Russo, James J.
; APPLICANT: Edelman, Isidore S.
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED
```

;; TITLE OF INVENTION: HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF
;; NUMBER OF SEQUENCES: 20
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Cooper & Dunham LLP
;; STREET: 1185 Avenue of the Americas
;; CITY: New York
;; STATE: New York
;; COUNTRY: U.S.A.
;; ZIP: 10036
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/770,379
;; FILING DATE:
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: White, John P.
;; REGISTRATION NUMBER: 28,678
;; REFERENCE/DOCKET NUMBER: 52342
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (212) 278-0400
;; TELEFAX: (212) 391-0525
;; INFORMATION FOR SEQ ID NO: 18:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 35100 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; MOLECULE TYPE: DNA (genomic)
US-08-770-379-18

Query Match 56.0%; Score 14; DB 2; Length 35100;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 gtgacgcgcagcat 24
|||||
Db 2058 TGCATGCCAGCAT 2045

RESULT 17
US-08-757-669A-18/c
;; Sequence 18, Application US/08757669A
;; Patent No. 6183751
;; GENERAL INFORMATION:
;; APPLICANT: Chang, Yuan
;; APPLICANT: Bohenzky, Roy A.
;; APPLICANT: Russo, James J.
;; APPLICANT: Edelman, Isidore S.
;; TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
;; TITLE OF INVENTION: SEQUENCES AND USES THEREOF
;; NUMBER OF SEQUENCES: 20
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Cooper & Dunham LLP
;; STREET: 1185 Avenue of the Americas
;; CITY: New York
;; STATE: New York
;; COUNTRY: U.S.A.
;; ZIP: 10036
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/757,669A
;; FILING DATE:
;; CLASSIFICATION: 424

;; ATTORNEY/AGENT INFORMATION:
;; NAME: White, John P.
;; REGISTRATION NUMBER: 28,678
;; REFERENCE/DOCKET NUMBER: 45185-F
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (212) 278-0400
;; TELEFAX: (212) 391-0525
;; INFORMATION FOR SEQ ID NO: 18:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 35100 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; MOLECULE TYPE: DNA (genomic)
US-08-757-669A-18

Query Match 56.0%; Score 14; DB 4; Length 35100;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 gtgacgcgcagcat 24
|||||
Db 2058 TGCATGCCAGCAT 2045

RESULT 18
US-09-060-756-726
;; Sequence 726, Application US/09060756
;; Patent No. 6183957
;; GENERAL INFORMATION:
;; APPLICANT: Cole, Stewart
;; APPLICANT: Buchrieser-Brosch, Roland
;; APPLICANT: Gordon, Stephen
;; APPLICANT: Billault, Alain
;; TITLE OF INVENTION: METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST FROM
;; TITLE OF INVENTION: THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED DNA
;; FILE REFERENCE: 3495-0169
;; CURRENT APPLICATION NUMBER: US/09/060,756
;; CURRENT FILING DATE: 1998-04-16
;; NUMBER OF SEQ ID NOS: 743
;; SOFTWARE: Patentin Ver. 2.0
;; SEQ ID NO 726
;; LENGTH: 304
;; TYPE: DNA
;; ORGANISM: Mycobacterium bovis
;; FEATURE:
;; NAME/KEY: unsure
;; LOCATION: (various positions within the sequence)
;; OTHER INFORMATION: applicants are uncertain of bases designated as "n"
US-09-060-756-726

Query Match 52.0%; Score 13; DB 4; Length 304;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 gtgacgcgcagc 22
|||||
Db 136 gtgacgcgcagc 148

RESULT 19
US-09-060-756-570
;; Sequence 570, Application US/09060756
;; Patent No. 6183957
;; GENERAL INFORMATION:
;; APPLICANT: Cole, Stewart
;; APPLICANT: Buchrieser-Brosch, Roland
;; APPLICANT: Gordon, Stephen
;; APPLICANT: Billault, Alain
;; TITLE OF INVENTION: METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST FROM

;; TITLE OF INVENTION: THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED DNA
;; FILE REFERENCE: 3495-0169
;; CURRENT APPLICATION NUMBER: US/09/060,756
;; CURRENT FILING DATE: 1998-04-16
;; NUMBER OF SEQ ID NOS: 743
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 570
;; LENGTH: 343
;; TYPE: DNA
;; ORGANISM: Mycobacterium tuberculosis
US-09-060-756-570

Query Match 52.0%; Score 13; DB 4; Length 343;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 gtgcacgcagc 22
|||||
DB 135 gtgcacgcagc 147

RESULT 20
US-08-640-978C-2/C
; Sequence 2, Application US/08640978C
; Patent No. 5932781
; GENERAL INFORMATION:
; APPLICANT: TAKANO, MITSUO
; APPLICANT: ONO, HISAYO
; APPLICANT: YAMADA, HIROYUKI
; APPLICANT: YAMATOYA, KAZUHIKO
; TITLE OF INVENTION: ECTOINE SYNTHASE GENE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH
; STREET: PO BOX 747
; CITY: FALLS CHURCH
; STATE: VA
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/640,978C
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: FARACI, C J
; REGISTRATION NUMBER: 32,350
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 411 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-640-978C-2

Query Match 52.0%; Score 13; DB 2; Length 411;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcgcagcaggtgc 13
|||||
DB 294 GCGCAGCAGGTGC 282

RESULT 21
US-08-640-978C-4/C
; Sequence 4, Application US/08640978C
; Patent No. 5932781
; GENERAL INFORMATION:
; APPLICANT: TAKANO, MITSUO
; APPLICANT: ONO, HISAYO
; APPLICANT: YAMADA, HIROYUKI
; APPLICANT: YAMATOYA, KAZUHIKO
; TITLE OF INVENTION: ECTOINE SYNTHASE GENE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH
; STREET: PO BOX 747
; CITY: FALLS CHURCH
; STATE: VA
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/640,978C
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: FARACI, C J
; REGISTRATION NUMBER: 32,350
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 414 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA to mRNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..411
US-08-640-978C-4

Query Match 52.0%; Score 13; DB 2; Length 414;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcgcagcaggtgc 13
|||||
DB 294 GCGCAGCAGGTGC 282

RESULT 22
US-08-836-075A-51
; Sequence 51, Application US/08836075A
; Patent No. 6180768
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; APPLICANT: STUYVER, LIEVEN
; TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
; TITLE OF INVENTION: AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC
; NUMBER OF SEQUENCES: 207
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ARNOLD, WHITE & DURKEE
; STREET: P.O. BOX 4433
; CITY: HOUSTON
; STATE: TEXAS

COUNTRY: USA
ZIP: 77210-4433
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Microsoft Word 6.0 / ASCII text output
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/836, 075A
FILING DATE: 21 Apr 1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP95/04155
FILING DATE: 23 Oct 1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 94870166.9
FILING DATE: 21 Oct 1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 95870076.7
FILING DATE: 28 Jun 1995
ATTORNEY/AGENT INFORMATION:
NAME: KAMMERER, PATRICIA A.
REGISTRATION NUMBER: 29,775
REFERENCE/DOCKET NUMBER: INNS:004
INFORMATION FOR SEQ ID NO: 51:
SEQUENCE CHARACTERISTICS:
LENGTH: 447 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-836-075A-51

Query Match 52.0%; Score 13; DB 4; Length 447;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 catgcgcagcatt 25
|||||
Db 406 CATGCAGCAGCATT 418

RESULT 23
US-08-998-416-978/C
Sequence 978, Application US/08998416
Patent No. 6239264
GENERAL INFORMATION:
APPLICANT: Philippsen, Peter
APPLICANT: Pohlmann, Rainer
APPLICANT: Steiner, Sabine
APPLICANT: Mohr, Christine
APPLICANT: Wendland, Jurgen
APPLICANT: Knechtel, Philipp
APPLICANT: Rebschunig, Corinne
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPTII
TITLE OF INVENTION: AND USES THEREOF
NUMBER OF SEQUENCES: 1152
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6239264rtls Corporation
STREET: 3054 Cornwalis Road
CITY: Research Triangle Park
STATE: No. 6239264th Carolina
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/998, 416

FILING DATE: 24-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 978:
SEQUENCE CHARACTERISTICS:
LENGTH: 476 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: PAG1602UP
US-08-998-416-978

Query Match 52.0%; Score 13; DB 4; Length 476;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ggcgcagcaggtgc 13
|||||
Db 328 GCGCAGCAGGTGC 316

RESULT 24
US-07-872-678A-11
Sequence 11, Application US/07872678A
Patent No. 5541060
GENERAL INFORMATION:
APPLICANT: Bell, Graeme, et al.
TITLE OF INVENTION: DETECTION OF EARLY-ONSET
TITLE OF INVENTION: NON-INSULIN-DEPENDENT DIABETES MELLITUS
NUMBER OF SEQUENCES: 48
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: Post Office Box 4433
CITY: Houston
STATE: Texas
COUNTRY: USA
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/872,678A
FILING DATE: 22-APRIL-1992
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Coughlin, Daniel F.
REGISTRATION NUMBER: 36,111
REFERENCE/DOCKET NUMBER: ARCD016
TELECOMMUNICATION INFORMATION:
TELEPHONE: 713-787-1400
TELEFAX: 713-789-2679
TELEX: 79-0924
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 1128 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)

US-07-872-678A-11

Query Match 52.0%; Score 13; DB 1; Length 1128;

Best Local Similarity 100.0%; Pred. No. 86;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 cgcagcaggtgca 14
|||||

Db 904 CGCAGCAGGTGCA 916

RESULT 25

US-08-428-414A-1

Sequence 1, Application US/08428414A

Patent No. 5912166

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF

TITLE OF INVENTION: LEISHMANIASIS

NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED and BERRY

STREET: 6300 Columbia Center, 701 Fifth Avenue

CITY: Seattle

STATE: Washington

COUNTRY: USA

ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/428,414A

CLASSIFICATION: 436

FILING DATE: 21-APR-1995

ATTORNEY/AGENT INFORMATION:

NAME: Kadlecek, Ann T.

REGISTRATION NUMBER: 39,244

REFERENCE/DOCKET NUMBER: 210121.407

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031

TELEX: 3723836 SEEDANBERRY

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1202 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

FEATURE:

NAME/KEY: CDS

LOCATION: 30..998

US-08-428-414A-1

Query Match 52.0%; Score 13; DB 2; Length 1202;

Best Local Similarity 100.0%; Pred. No. 85;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 cgcagcaggtgca 14
|||||

Db 133 CGCAGCAGGTGCA 145

RESULT 26

US-09-319-892-1

Sequence 1, Application US/09319892

Patent No. 6177616

GENERAL INFORMATION:

APPLICANT: BARTSCH, Klaus

APPLICANT: KRIETE, Guido

APPLICANT: BROER, Inge

APPLICANT: PUHLER, Alfred

TITLE OF INVENTION: NOVEL GENES CODING FOR AMINO ACID DEACETYLASES WITH

TITLE OF INVENTION: SPECIFICITY FOR N-ACETYL-L-PHOSPHOTRITHICIN, THEIR

TITLE OF INVENTION: ISOLATION AND THEIR USE

FILE REFERENCE: 514412-2005

CURRENT APPLICATION NUMBER: US/09/319,892

CURRENT FILING DATE: 1999-06-14

EARLIER APPLICATION NUMBER: PCT/EP97/06755

EARLIER FILING DATE: 1997-12-03

EARLIER APPLICATION NUMBER: 19652284.6

EARLIER FILING DATE: 1996-12-16

NUMBER OF SEQ ID NOS: 4

SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 1

LENGTH: 1365

TYPE: DNA

ORGANISM: Stenotrophomonas maltophilia

US-09-319-892-1

Query Match 52.0%; Score 13; DB 4; Length 1365;

Best Local Similarity 100.0%; Pred. No. 85;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcagcaggtgc 13
|||||

Db 1139 ggcagcaggtgc 1151

RESULT 27

US-07-862-588B-6

Sequence 6, Application US/07862588B

Patent No. 5916796

GENERAL INFORMATION:

APPLICANT: Joergensen, Per Linnaa

APPLICANT: Sch lein, Martin

APPLICANT: Hansen, Christian

TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity

NUMBER OF SEQUENCES: 7

CORRESPONDENCE ADDRESS:

ADDRESSEE: No. 59167960 No. 5916796disk of No. 5916796th America, Inc.

STREET: 405 Lexington Avenue, 62nd floor

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10017

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/862,588B

CLASSIFICATION: 435

FILING DATE: 19920727

PRIOR APPLICATION DATA:

APPLICATION NUMBER: DK 164/90

FILING DATE: 19-JAN-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/DK91/00013

FILING DATE: 18-JAN-1991

ATTORNEY/AGENT INFORMATION:

NAME: zelson, Steve T. / lambiris, Elias J.

REGISTRATION NUMBER: 30,335 / 33,728

REFERENCE/DOCKET NUMBER: 3425.204-US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212 867 0123

TELEFAX: 212 867 0298

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 1624 base pairs

TYPE: NUCLEIC ACID

```
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: Bacillus lautus
STRAIN: NCIMB 40250
FEATURE:
NAME/KEY: CDS
LOCATION: 30..1607
OTHER INFORMATION:
US-07-862-588B-6

Query Match          52.0%; Score 13; DB 2; Length 1624;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 caggtgcatgcc 19
    |||
Db 934 CAGGTGCATGCC 946

RESULT 28
US-07-862-588B-5
; Sequence 5, Application US/07862588B
; Patent No. 5916796
; GENERAL INFORMATION:
; APPLICANT: Joergensen, Per Linna
; APPLICANT: Sch lein, Martin
; APPLICANT: Hansen, Christian
; TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESS: NO. 59167960 No. 5916796disk of No. 5916796th America, Inc.
; STREET: 405 Lexington Avenue, 62nd floor
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/862,588B
; FILING DATE: 19920727
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 164/90
; FILING DATE: 19-JAN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DK91/00013
; FILING DATE: 18-JAN-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Zelson, Steve T. / Lambiris, Elias J.
; REGISTRATION NUMBER: 30,335 / 33,728
; REFERENCE/DOCKET NUMBER: 3425.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 867 0123
; TELEFAX: 212 867 0298
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1775 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Bacillus lautus
; STRAIN: NCIMB.40250
```

```
FEATURE:
NAME/KEY: CDS
LOCATION: 30..(1625.1775)
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /partial
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /transl_except= (pos: 1446 .. 1458, aa:
; OTHER INFORMATION: OTR)
US-07-862-588B-5

Query Match          52.0%; Score 13; DB 2; Length 1775;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 caggtgcatgcc 19
    |||
Db 934 CAGGTGCATGCC 946

RESULT 29
US-09-258-373-21/C
; Sequence 21, Application US/09258373
; Patent No. 6150110
; GENERAL INFORMATION:
; APPLICANT: Fletcher, Jonathan A.
; APPLICANT: Xiao, Sheng
; TITLE OF INVENTION: HMG(Y)-LAMA4* FUSION ONCOGENE,
; FILE REFERENCE: B0801/7135/ERP
; CURRENT APPLICATION NUMBER: US/09/258,373
; CURRENT FILING DATE: 1999-02-26
; EARLIER APPLICATION NUMBER: 60/076,401
; EARLIER FILING DATE: 1998-02-28
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 21
; LENGTH: 1875
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-09-258-373-21

Query Match          52.0%; Score 13; DB 3; Length 1875;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 ccgacgcaggtgca 14
    |||
Db 838 CGCAGCAGGTGCA 826

RESULT 30
US-09-017-706-3
; Sequence 3, Application US/09017706A
; Patent No. 6087147
; GENERAL INFORMATION:
; APPLICANT: ITO, YOSHIFUMI
; TITLE OF INVENTION: A-AMYLASE GENE HAVING ABILITY FOR HIGHLY PRODUCING
; TITLE OF INVENTION: MALTOPEPTASE, VECTOR CONTAINING SAID GENE AND
; FILE REFERENCE: 8361-0003-0
; CURRENT APPLICATION NUMBER: US/09/017,706A
; CURRENT FILING DATE: 1998-02-05
; EARLIER APPLICATION NUMBER: JP 305071/1997
; EARLIER FILING DATE: 1997-10-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1981
; TYPE: DNA
; ORGANISM: Pseudomonas sp., Strain KO-8940
; FEATURE:
```

```
; NAME/KEY: CDS
; LOCATION: (7)..(1848)
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: (85)..(1848)
; OTHER INFORMATION: METHOD OF DETERMINING CHARACTERISTIC: E
; FEATURE:
; OTHER INFORMATION: MUTATED GENOMIC DNA
; FEATURE:
; OTHER INFORMATION: PLASMID POS3410F57
; US-09-017-706-3
```

```
Query Match          52.0%; Score 13; DB 3; Length 1981;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      2  ccgacagcagtgca 14
        |||||||
Db      1530  ccgacagcagtgca 1542
```

```
RESULT 31
US-09-017-706-4
; Sequence 4, Application US/09017706A
; Patent No. 6087147
; GENERAL INFORMATION:
; APPLICANT: ITO, YOSHIFUMI
; TITLE OF INVENTION: A-AMYLAASE GENE HAVING ABILITY FOR HIGHLY PRODUCING
; TITLE OF INVENTION: MALTOPEPTASE, VECTOR CONTAINING SAID GENE AND
; FILE REFERENCE: 8361-0003-0
; CURRENT APPLICATION NUMBER: US/09/017,706A
; EARLIER FILING DATE: 1998-02-05
; EARLIER APPLICATION NUMBER: JP 305071/1997
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 1981
; TYPE: DNA
; ORGANISM: Pseudomonas sp., Strain KO-8940
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (7)..(1848)
; NAME/KEY: mat_peptide
; LOCATION: (85)..(1848)
; OTHER INFORMATION: METHOD OF DETERMINING CHARACTERISTIC: E
; FEATURE:
; OTHER INFORMATION: MUTATED GENOMIC DNA
; FEATURE:
; OTHER INFORMATION: PLASMID POS3410H57
; US-09-017-706-4
```

```
Query Match          52.0%; Score 13; DB 3; Length 1981;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      2  ccgacagcagtgca 14
        |||||||
Db      1530  ccgacagcagtgca 1542
```

```
RESULT 32
US-09-017-706-5
; Sequence 5, Application US/09017706A
; Patent No. 6087147
; GENERAL INFORMATION:
; APPLICANT: ITO, YOSHIFUMI
; TITLE OF INVENTION: A-AMYLAASE GENE HAVING ABILITY FOR HIGHLY PRODUCING
; TITLE OF INVENTION: MALTOPEPTASE, VECTOR CONTAINING SAID GENE AND
```

```
; TITLE OF INVENTION: TRANSFORMANT
; FILE REFERENCE: 8361-0003-0
; CURRENT APPLICATION NUMBER: US/09/017,706A
; CURRENT FILING DATE: 1998-02-05
; EARLIER APPLICATION NUMBER: JP 305071/1997
; EARLIER FILING DATE: 1997-10-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 1981
; TYPE: DNA
; ORGANISM: Pseudomonas sp., Strain KO-8940
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (7)..(1848)
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: (85)..(1848)
; OTHER INFORMATION: METHOD OF DETERMINING CHARACTERISTIC: E
; FEATURE:
; OTHER INFORMATION: MUTATED GENOMIC DNA
; FEATURE:
; OTHER INFORMATION: PLASMID POS3410OL57
; US-09-017-706-5
```

```
Query Match          52.0%; Score 13; DB 3; Length 1981;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      2  ccgacagcagtgca 14
        |||||||
Db      1530  ccgacagcagtgca 1542
```

```
RESULT 33
US-09-017-706-6
; Sequence 6, Application US/09017706A
; Patent No. 6087147
; GENERAL INFORMATION:
; APPLICANT: ITO, YOSHIFUMI
; TITLE OF INVENTION: A-AMYLAASE GENE HAVING ABILITY FOR HIGHLY PRODUCING
; TITLE OF INVENTION: MALTOPEPTASE, VECTOR CONTAINING SAID GENE AND
; FILE REFERENCE: 8361-0003-0
; CURRENT APPLICATION NUMBER: US/09/017,706A
; EARLIER FILING DATE: 1998-02-05
; EARLIER APPLICATION NUMBER: JP 305071/1997
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 1981
; TYPE: DNA
; ORGANISM: Pseudomonas sp., Strain KO-8940
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (7)..(1848)
; NAME/KEY: mat_peptide
; LOCATION: (85)..(1848)
; FEATURE:
; OTHER INFORMATION: MUTATED GENOMIC DNA
; FEATURE:
; OTHER INFORMATION: PLASMID: POS3410F139
; US-09-017-706-6
```

```
Query Match          52.0%; Score 13; DB 3; Length 1981;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      2  ccgacagcagtgca 14
```

Db 1530 cgcagcaggtgca 1542

RESULT 34

US-09-017-706-7
 ; Sequence 7, Application US/09017706A
 ; Patent No. 6087147
 ; GENERAL INFORMATION:
 ; APPLICANT: ITO, YOSHIFUMI
 ; TITLE OF INVENTION: A-AMYLASE GENE HAVING ABILITY FOR HIGHLY PRODUCING
 ; TITLE OF INVENTION: MALTOPENLOSE, VECTOR CONTAINING SAID GENE AND
 ; FILE REFERENCE: 8361-0003-0
 ; CURRENT APPLICATION NUMBER: US/09/017,706A
 ; CURRENT FILING DATE: 1998-02-05
 ; EARLIER APPLICATION NUMBER: JP 305071/1997
 ; EARLIER FILING DATE: 1997-10-21
 ; NUMBER OF SEQ ID NOS: 14
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 7
 ; LENGTH: 1981
 ; TYPE: DNA
 ; ORGANISM: Pseudomonas sp., Strain KO-8940
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (7)..(1848)
 ; FEATURE:
 ; NAME/KEY: mat_peptide
 ; LOCATION: (85)..(1848)
 ; OTHER INFORMATION: METHOD OF DETERMINING CHARACTERISTIC: E
 ; FEATURE:
 ; OTHER INFORMATION: MUTATED GENOMIC DNA
 ; OTHER INFORMATION: PLASMID: POS3410H139
 ; US-09-017-706-7

Query Match 52.0%; Score 13; DB 3; Length 1981;
 Best Local Similarity 100.0%; Pred. No. 85;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 cgcagcaggtgca 14
 Db 1530 cgcagcaggtgca 1542

RESULT 35
 US-09-017-706-8
 ; Sequence 8, Application US/09017706A
 ; Patent No. 6087147
 ; GENERAL INFORMATION:
 ; APPLICANT: ITO, YOSHIFUMI
 ; TITLE OF INVENTION: A-AMYLASE GENE HAVING ABILITY FOR HIGHLY PRODUCING
 ; TITLE OF INVENTION: MALTOPENLOSE, VECTOR CONTAINING SAID GENE AND
 ; FILE REFERENCE: 8361-0003-0
 ; CURRENT APPLICATION NUMBER: US/09/017,706A
 ; CURRENT FILING DATE: 1998-02-05
 ; EARLIER APPLICATION NUMBER: JP 305071/1997
 ; EARLIER FILING DATE: 1997-10-21
 ; NUMBER OF SEQ ID NOS: 14
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 8
 ; LENGTH: 1981
 ; TYPE: DNA
 ; ORGANISM: Pseudomonas sp., Strain KO-8940
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (7)..(1848)
 ; FEATURE:
 ; NAME/KEY: mat_peptide
 ; LOCATION: (85)..(1848)
 ; LOCATION: (85)..(1848)

FEATURE:
 ; OTHER INFORMATION: MUTATED GENOMIC DNA
 ; FEATURE:
 ; OTHER INFORMATION: PLASMID: POS3410L139
 ; US-09-017-706-8

Query Match 52.0%; Score 13; DB 3; Length 1981;
 Best Local Similarity 100.0%; Pred. No. 85;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 cgcagcaggtgca 14
 Db 1530 cgcagcaggtgca 1542

RESULT 36
 US-08-317-522A-8/C
 ; Sequence 8, Application US/08317522A
 ; Patent No. 559918
 ; GENERAL INFORMATION:
 ; APPLICANT: Fukuda, Michiko N.
 ; TITLE OF INVENTION: Trophinin and Trophinin-Assisting
 ; TITLE OF INVENTION: Proteins
 ; NUMBER OF SEQUENCES: 13
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Campbell and Flores
 ; STREET: 4370 La Jolla Village Drive, Suite 700
 ; CITY: San Diego
 ; STATE: California
 ; COUNTRY: USA
 ; ZIP: 92122
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/317,522A
 ; FILING DATE: 04-OCT-1994
 ; CLASSIFICATION: 536
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Campbell, Cathryn A.
 ; REGISTRATION NUMBER: 31,815
 ; REFERENCE/DOCKET NUMBER: P-LA 9991
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (619) 535-9001
 ; TELEFAX: (619) 535-8949
 ; INFORMATION FOR SEQ ID NO: 8:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 223 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 199..2223
 ; US-08-317-522A-8

Query Match 52.0%; Score 13; DB 1; Length 2223;
 Best Local Similarity 100.0%; Pred. No. 85;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ggcgacaggtgc 13
 Db 2118 GCGCAGCAGGTGC 2106

RESULT 37
 US-08-439-818A-8/C
 ; Sequence 8, Application US/08439818A
 ; Patent No. 5654145

GENERAL INFORMATION:
APPLICANT: Fukuda, Michiko N.
TITLE OF INVENTION: Trophinin and Trophinin-Assisting
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/439,818A
FILING DATE: 12-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/317,522
FILING DATE: 04-OCT-1994
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LA 1563
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 2223 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 199..2223
US-08-439-818A-8

Query Match 52.0%; Score 13; DB 1; Length 2223;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gccgacgaggtgc 13
|||||
Db 2118 gccgacgaggtgc 2106

RESULT 38
US-08-751-965-8/c
Sequence 8, Application US/08751965
Patent No. 5858360
GENERAL INFORMATION:
APPLICANT: Fukuda, Michiko N.
TITLE OF INVENTION: Trophinin and Trophinin-Assisting
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/751,965
FILING DATE: Herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/439,818
FILING DATE: 12-MAY-1995
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LA 2252
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 2223 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 199..2223
US-08-751-965-8

Query Match 52.0%; Score 13; DB 2; Length 2223;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gccgacgaggtgc 13
|||||
Db 2118 gccgacgaggtgc 2106

RESULT 39
US-08-738-975-8/c
Sequence 8, Application US/08738975
Patent No. 5880267
GENERAL INFORMATION:
APPLICANT: Fukuda, Michiko N.
TITLE OF INVENTION: Trophinin and Trophinin-Assisting
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/738,975
FILING DATE: herewith
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/439,818
FILING DATE: 05-Dec-1995
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LA 2251
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 2223 base pairs

TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 199..2223
US-08-738-975-8

Query Match 52.0%; Score 13; DB 2; Length 2223;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggcagcaggtgc 13
|||||
Db 2118 GCGCAGCAGGTGC 2106

RESULT 40
US-08-728-626-8/c
Sequence 8, Application US/08728626
Patent No. 5910451
GENERAL INFORMATION:
APPLICANT: Fukuda, Michiko N.
TITLE OF INVENTION: Trophinin and Trophinin-Assisting
PROTEINS
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/728.626
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/439,818
FILING DATE: 12-MAY-1995
APPLICATION NUMBER: US 08/317,522
FILING DATE: 04-OCT-1994
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LA 1563
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 2223 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 199..2223
US-08-728-626-8

Query Match 52.0%; Score 13; DB 2; Length 2223;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ggcagcaggtgc 13
|||||

Db 2118 GCGCAGCAGGTGC 2106

RESULT 41
US-08-808-599A-8/c
Sequence 8, Application US/08808599A
Patent No. 611089
GENERAL INFORMATION:
APPLICANT: Fukuda, Michiko N.
TITLE OF INVENTION: Trophinin, Trophinin-Assisting
PROTEINS and Methods to Inhibit Implantation
NUMBER OF SEQUENCES: 41
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell & Flores LLP
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/808.599A
FILING DATE: 28-FEB-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/317,522
FILING DATE: 04-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/439,818
FILING DATE: 12-MAY-1995
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LA 2256
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 2223 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 199..2223
US-08-808-599A-8

Query Match 52.0%; Score 13; DB 3; Length 2223;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ggcagcaggtgc 13
|||||
Db 2118 GCGCAGCAGGTGC 2106

RESULT 42
US-08-584-226-1/c
Sequence 1, Application US/08584226
Patent No. 5798240
GENERAL INFORMATION:
APPLICANT: Martinis, Susan A.
APPLICANT: Sassanfar, Mandana
APPLICANT: Kim, Sunghoon
APPLICANT: Lee, Sang Ho
APPLICANT: Schimmel, Paul R.
TITLE OF INVENTION: RECOMBINANT MYCOBACTERIAL METHIONYL-tRNA

TITLE OF INVENTION: SYNTHETASE GENES, TESTER STRAINS AND ASSAYS
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Millitia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: USA
ZIP: 02173-4799
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/584,226
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/305,766
FILING DATE: 13-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Brook, David E.
REGISTRATION NUMBER: 22,592
REFERENCE/DOCKET NUMBER: CPI94-052
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-9540
TELEFAX: 617-861-6240
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2290 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1563
US-08-584-226-1

Query Match 52.0%; Score 13; DB 1; Length 2290;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 caggtgcatgcgc 19
|||||
Db 1233 CAGGTGCATGCC 1221

RESULT 43
US-08-941-445A-16/C
Sequence 16, Application US/08941445A
Patent No. 6107060
GENERAL INFORMATION:
APPLICANT: Keeling, Peter
APPLICANT: Guan, Hanning
TITLE OF INVENTION: Starch Encapsulation
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
STREET: 5370 Manhattan Circle
CITY: Boulder
STATE: CO
COUNTRY: US
ZIP: 80303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/941,445A

FILING DATE: 30-SEP-1997
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/026,855
FILING DATE: 30-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Winner, Ellen P
REGISTRATION NUMBER: 28,547
REFERENCE/DOCKET NUMBER: 89-97
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 499-8080
TELEFAX: (303) 499-8089
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 2763 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: not relevant
MOLECULE TYPE: mRNA
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Zea mays
FEATURE:
NAME/KEY: translt-peptide
LOCATION: 2..190
FEATURE:
NAME/KEY: mat-peptide
LOCATION: 191..2467
FEATURE:
NAME/KEY: CDS
LOCATION: 2..2470
US-08-941-445A-16

Query Match 52.0%; Score 13; DB 3; Length 2763;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcgcaggtgc 13
|||||
Db 443 GCGCAGCAGGTGC 431

RESULT 44
US-07-872-678A-36
Sequence 36, Application US/07872678A
Patent No. 5541060
GENERAL INFORMATION:
APPLICANT: Bell, Graeme, et al.
TITLE OF INVENTION: DETECTION OF EARLY-ONSET
NUMBER OF SEQUENCES: 48
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: Post Office Box 4433
CITY: Houston
STATE: Texas
COUNTRY: USA
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/872,678A
FILING DATE: 22-APRIL-1992
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Coughlin, Daniel F.
REGISTRATION NUMBER: 36,111
REFERENCE/DOCKET NUMBER: ARCD016
TELECOMMUNICATION INFORMATION:

TELEPHONE: 713-787-1400
 TELEFAX: 713-789-2679
 TELE: 79-0924
 INFORMATION FOR SEQ ID NO: 36:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 3618 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 US-07-872-678A-36

Query Match 52.0%; Score 13; DB 1; Length 3618;
 Best Local Similarity 100.0%; Pred. No. 85;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

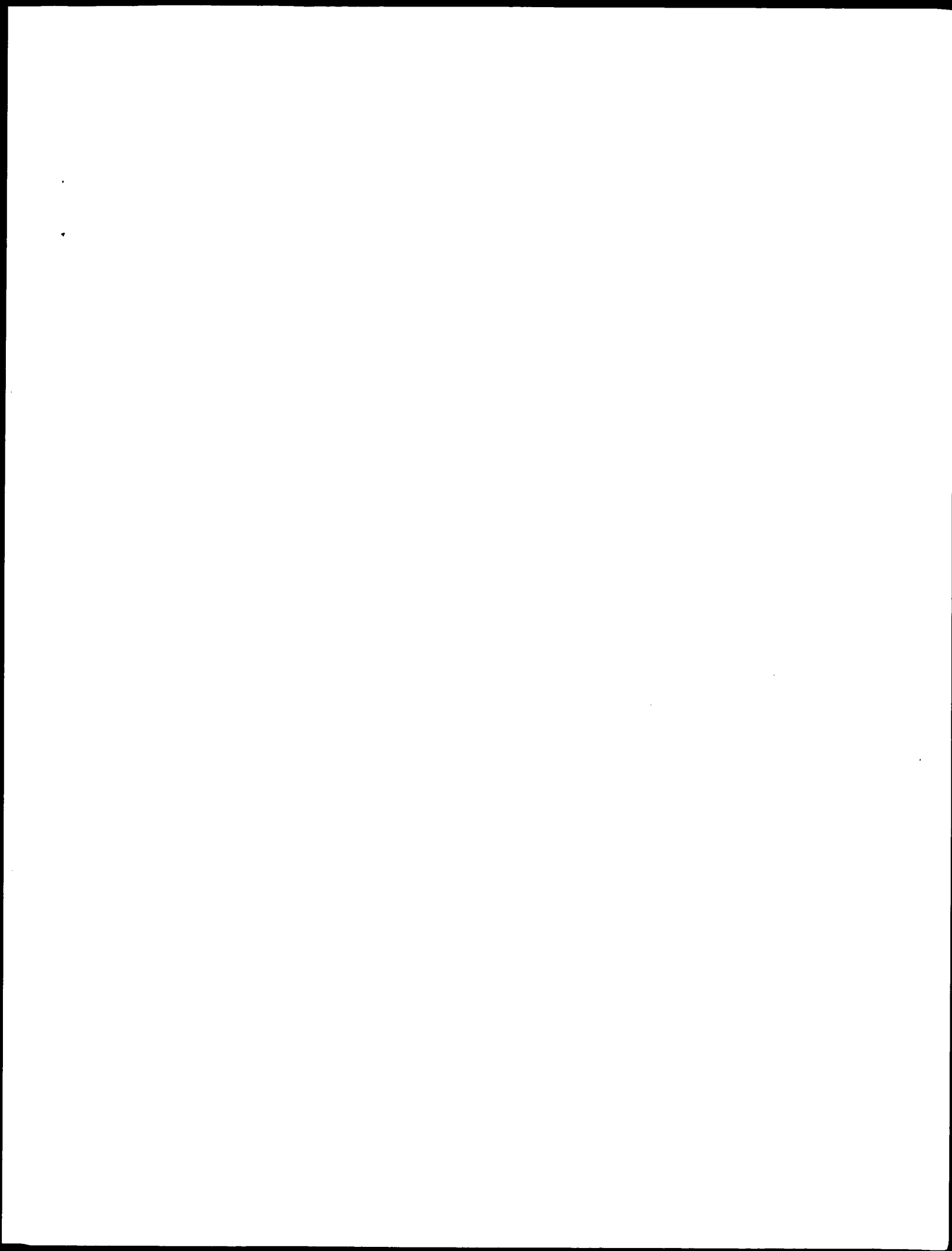
QY 2 ccgcagcaggtgca 14
 |||
 Db 2446 ccgcagcaggtgca 2458

RESULT 45
 US-08-386-727-3/c
 Sequence 3, Application US/08386727
 Patent No. 5792647
 GENERAL INFORMATION:
 APPLICANT: ROSEMAN, SAUL
 APPLICANT: BASSLER, BONNIE
 APPLICANT: KEYHANI, NEMAT O.
 APPLICANT: CHITLARU, EDITH
 APPLICANT: ROWE, CHRIS
 APPLICANT: YU, CHARLES
 TITLE OF INVENTION: BACTERIAL CATABOLISM OF CHITIN
 NUMBER OF SEQUENCES: 8
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: CUSHMAN, DARBY & CUSHMAN
 STREET: 1100 NEW YORK AVENUE, N.W.
 CITY: WASHINGTON
 STATE: D.C.
 COUNTRY: USA
 ZIP: 20005
 COMPUTER READABLE FORM:
 MEDIUM TYPE: FLOPPY disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentln Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/386,727
 FILING DATE:
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: HOBBS, ANN S.
 REGISTRATION NUMBER: 36,830
 REFERENCE/DOCKET NUMBER: 4130/206916
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 202-861-3000
 TELEFAX: 202-822-0944
 TELEX: 6714627 CUSH
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 3670 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 US-08-386-727-3

Query Match 52.0%; Score 13; DB 1; Length 3670;
 Best Local Similarity 100.0%; Pred. No. 85;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gccgcagcaggtgc 13
 |||
 Db 2248 gccgcagcaggtgc 2236

Search completed: October 9, 2001, 15:55:32
 Job time: 13208 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 15:15:28 ; Search time 5323.87 Seconds
(without alignments)
44.389 Million cell updates/sec

Title: US-09-396-196f-5

Perfect score: 25
Sequence: 1 gcgcagcagtgcatgccacgacatt 25

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 10228115 seqs, 4726426750 residues

Word size : 9

Total number of hits satisfying chosen parameters: 520956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: gb_est41:*
42: gb_est42:*
43: gb_est43:*

44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_esthum29:*
64: em_esthum30:*
65: em_esthum31:*
66: em_esthum32:*
67: em_esthum33:*
68: em_esthum34:*
69: em_esthum35:*
70: em_esthum36:*
71: em_esthum37:*
72: em_esthum38:*
73: em_esthum39:*
74: em_esthum40:*
75: em_esthum41:*
76: em_esthum42:*
77: em_esthum43:*
78: em_esthum44:*
79: em_esthum45:*
80: em_esthum46:*
81: em_esthum47:*
82: em_esthum48:*
83: em_esthum49:*
84: em_esthum50:*
85: em_esthum51:*
86: em_esthum52:*
87: em_esthum53:*
88: em_esthum54:*
89: em_esthum55:*
90: em_esthum56:*
91: em_esthum57:*
92: em_esthum58:*
93: em_esthum59:*
94: em_esthum60:*
95: em_esthum61:*
96: em_esthum62:*
97: em_esthum63:*
98: em_esthum64:*
99: em_esthum65:*
100: em_esthum66:*
101: em_esthum67:*
102: em_esthum68:*
103: em_esthum69:*
104: em_esthum70:*
105: em_esthum71:*
106: em_esthum72:*
107: em_esthum73:*
108: em_esthum74:*
109: em_esthum75:*
110: em_esthum76:*
111: em_esthum77:*
112: em_esthum78:*
113: em_esthum79:*
114: em_esthum80:*
115: em_esthum81:*
116: em_esthum82:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

190: gb_est110:*
191: gb_est111:*
192: gb_hic:*
193: em_gss_fun:*
194: em_gss_hum1:*
195: em_gss_hum2:*
196: em_gss_hum3:*
197: em_gss_hum4:*
198: em_gss_hum5:*
199: em_gss_hum6:*
200: em_gss_hum7:*
201: em_gss_hum8:*
202: em_gss_hum9:*
203: em_gss_inv1:*
204: em_gss_inv2:*
205: em_gss_inv3:*
206: em_gss_inv4:*
207: em_gss_inv5:*
208: em_gss_inv6:*
209: em_gss_inv7:*
210: em_gss_inv8:*
211: em_gss_inv9:*
212: em_gss_inv10:*
213: em_gss_inv11:*
214: em_gss_inv12:*
215: em_gss_inv13:*
216: em_gss_inv14:*
217: em_gss_inv15:*
218: em_gss_inv16:*
219: em_gss_inv17:*
220: em_gss_inv18:*
221: em_gss_inv19:*
222: em_gss_inv20:*
223: em_gss_inv21:*
224: em_gss_inv22:*
225: em_gss_inv23:*
226: em_gss_inv24:*
227: em_gss_inv25:*
228: em_gss_inv26:*
229: em_gss_inv27:*
230: em_gss_inv28:*
231: em_gss_inv29:*
232: em_gss_inv30:*
233: em_gss_inv31:*
234: em_gss_inv32:*
235: em_gss_inv33:*
236: em_gss_inv34:*
237: em_gss_inv35:*
238: em_gss_inv36:*
239: em_gss_inv37:*
240: em_gss_inv38:*
241: em_gss_inv39:*
242: em_gss_inv40:*
243: em_gss_inv41:*
244: em_gss_inv42:*
245: em_gss_inv43:*
246: em_gss_inv44:*
247: em_gss_inv45:*
248: em_gss_inv46:*
249: em_gss_inv47:*
250: em_gss_inv48:*
251: em_gss_inv49:*
252: em_gss_inv50:*
253: em_gss_inv51:*
254: em_gss_inv52:*
255: em_gss_inv53:*
256: em_gss_inv54:*
257: em_gss_inv55:*
258: em_gss_inv56:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

JOURNAL
MEDLINE
COMMENT

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20020663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001

Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?lcr-CM1<2-CM1-NT0270-
211200-661-g03<3=2000-12-21<4=1)

Seq primer: puc 18 forward
High quality sequence start: 10
High quality sequence stop: 376.

FEATURES
source

1..376
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_id="NT0270"
/dev_stage="Adult"

/note="Organ: nervous_tumor; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

BASE COUNT
ORIGIN
77 a 104 c 99 g 96 t

Query Match 64.0%; Score 16; DB 171; Length 376;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 9 ggtgcacgcacgcat 24
|||||
Db 329 ggtgcacgcacgcat 314

RESULT 3
D24778 399 bp mRNA EST 08-JUL-1999
LOCUS R1CR2545A Rice root Oryza sativa cDNA clone R2545_1A, mRNA
DEFINITION D24778
ACCESSION D24778.1 GI:428626
KEYWORDS EST.
SOURCE Oryza sativa.
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliopsida; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
1 (bases 1 to 399)
Minobe,Y. and Sasaki,T.
Rice cDNA from root
Unpublished (1995)
Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program
2-1-2 Kannondai, Tsukuba
Ibaraki,
Japan 305
Tel: 0298-38-7441
Fax: 0298-38-7468
Email: tsasaki@agr.affrc.go.jp
PRODUCT "RCP".

FEATURES
source
1..399
/organism="Oryza sativa"

/strain="Nipponbare, sub-species Japonica"
/db_xref="taxon:4530"
/clone_id="R2545_1A"
/clone_lib="Rice root"
/note="Prepared from seedling root."

BASE COUNT 114 a 70 c 102 g 112 t 1 others
ORIGIN
Query Match 64.0%; Score 16; DB 156; Length 399;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 9 ggtgcacgcacgcat 24
|||||
Db 251 ggtgcacgcacgcat 266

RESULT 4
A0852582 479 bp DNA GSS 13-APR-2001
LOCUS LMAJFV1_lm74c04.y1 Leishmania major FV1 random genomic library
DEFINITION LMAJFV1_lm74c04.y1 Leishmania major genomic clone LMAJFV1_lm74c04 5', DNA sequence.
ACCESSION A0852582
VERSION A0852582.1 GI:6118907
KEYWORDS GSS.
SOURCE Leishmania major.
ORGANISM Leishmania major
Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
Leishmania.
1 (bases 1 to 479)
Akopyants,N.S., Clifton,S.W., Martin,J., Pape,D., Wylie,T., Li,L.,
Kissinger,J., Roos,D.S., Marra,M., Hillier,L., Chinwalla,A.,
Blistain,A., Schmitt,A., Peterson,B., Theisling,B., Ritter,E., Ronko,
I., Bennett,J., Cole,R., Underwood,K., Cardenas,M., Gibbons,M.,
Harvey,N., McCann,R., Tsagaris,I., Williams,T., Jackson,Y.,
Bowers,Y., Swaller,T., Waterston,R., Wilson,R. and Beverley,S.M.
A survey of the Leishmania major Friedlin strain VI genome by
shotgun sequencing: a resource for DNA
Mol. Biochem. Parasitol. 113 (2), 337-340 (2001)
Contact: Akopyants, NS / Beverley, SM
WashU Leishmania Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewartson.wustl.edu

Library construction: Natalia S. Akopyants, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
If using this information please cite:
N.S. Akopyants and S.M. Beverley 'A survey of the Leishmania major
Friedlin strain VI genome by shotgun sequencing' and the Washington
University Genome Sequencing Center for information on obtaining
clone material please contact: Natalia S. Akopyants Ph.D.
(natalia@borcim.wustl.edu) and/or Stephen M. Beverley Ph.D.
(beverley@borcim.wustl.edu)
Seq primer: -40RP from Gibco
Class: shotgun
High quality sequence stop: 372.

FEATURES
source

1..479
/organism="Leishmania major"
/strain="Friedlin strain VI"
/db_xref="taxon:5664"
/clone="LMAJFV1_lm74c04"
/clone_lib="Leishmania major FV1 random genomic library"
/lab_host="TOP10 (Invitrogen)"
/note="Vector: pZero-2 (Invitrogen); Site_1: EcoRV;
Genomic DNA was isolated from stationary phase cells. For
this library, DNA was sheared to give a tight size
distribution of 1-1.5kb fragments, blunt-ended with T4 DNA
polymerase, dephosphorylated with Shrimp Alkaline
Phosphatase and ligated into pZero-2 vector's EcoRV site."

BASE COUNT 98 a 155 c 147 g 79 t
 ORIGIN

Query Match 64.0%; Score 16; DB 234; Length 479;
 Best Local Similarity 100.0%; Pred. No. 52;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 cagcagtcgcatgcc 19
 |||
 Db 47 CAGCAGTCGATGCC 62

RESULT 5
 P968R/c 548 bp DNA GSS 25-JUL-2000
 LOCUS
 DEFINITION Leishmania major Friedlin PAC P968 right end-sequence, genomic survey sequence.
 ACCESSION AL390674
 KEYWORDS GSS.
 SOURCE Leishmania major.
 ORGANISM Leishmania major.
 Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania
 1 (bases 1 to 548)
 Ivens,A.C., Lewis,S.M., Bagherzadeh,A., Zhang,L., Chan,H.M. and Smith,D.F.
 TITLE A physical map of the Leishmania major Friedlin genome
 JOURNAL Genome Res. 8 (2), 135-145 (1998)
 MEDLINE 98146435
 REFERENCE 2 (bases 1 to 548)
 AUTHORS Taylor,R.G., Huckle,E.E.J., Ivens,A.C., Rajandream,M.A. and Barrell,B.G.
 TITLE Direct Submission
 JOURNAL Submitted (24-JUL-2000) Leishmania major Friedlin genome sequencing project, Sanger Centre, The Wellcome Trust Genome Campuss, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and allcats@sanger.ac.uk
 COMMENT see http://www.ebi.ac.uk/parasites/leish.html
 DETAILS of Leishmania sequencing at the Sanger Centre are available at http://www.sanger.ac.uk/Projects/L.major/
 The primer sequence can be obtained from allcats@sanger.ac.uk.

FEATURES
 source
 1..548
 /organism="Leishmania major"
 /strain="Friedlin"
 /db_xref="taxon:5664"
 /clone="PAC P968"

BASE COUNT 86 a 167 c 163 g 132 t
 ORIGIN

Query Match 64.0%; Score 16; DB 258; Length 548;
 Best Local Similarity 100.0%; Pred. No. 52;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 gtgcagtcgcatgcc 25
 |||
 Db 541 GTGCATGCCGATCAT 526

RESULT 6
 BF864751 761 bp mRNA EST 19-JAN-2001
 LOCUS 963054E05.y1 C. reinhardtii CC-1690, Stress condition I, normalized
 DEFINITION Lambda zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.
 ACCESSION BF864751
 KEYWORDS EST.
 SOURCE Chlamydomonas reinhardtii.
 ORGANISM Chlamydomonas reinhardtii
 Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;

REFERENCE
 AUTHORS Grossman,A., Davies,J., Federspiel,N., Harris,E., Hauser,C., Lefebvre,P., McDermott,J.P., Strager,J., Sillow,C. and Stern,D.
 TITLE Analyses of the Chlamydomonas reinhardtii Genome: A Model, Unicellular System for Analyzing Gene Function and Regulation in Vascular Plants: project phase 3
 JOURNAL Unpublished (2000)
 COMMENT Contact: Charles Hauser
 DCMB Box 91000
 Duke University
 Durham, NC 27708-1000
 Tel: 919 613 8159
 Fax: 919 613 8177
 Email: chauser@duke.edu

FEATURES
 source
 1..761
 /organism="Chlamydomonas reinhardtii"
 /strain="CC-1690 wild type mt+ 21gr"
 /db_xref="taxon:3055"
 /clone_lib="C. reinhardtii CC-1690, Stress condition I, normalized, Lambda zap II"
 /note="Vector: Bluescript II SK-; Site_1: EcoRI; Site_2: XhoI; This library, constructed by John Davies and Jeffrey McDermott, combines cDNAs from CC-1690 cells grown to mid-log phase in TAP-N (30 min, 1hr, 4hr), TAP-S (30 min, 1hr, 4hr), TAP-P (4hr, 12hr, 24hr), NO3 to NH4 (30min, 1hr, 4hr) and NH4 to NO3 (30min, 1hr, 4hr). PolyA mRNA was purified from each sample, pooled and cDNA synthesized. The cDNA was directionally cloned into lambda zap II (Stratagene) in the EcoRI (5') and XhoI (3') sites. Bluescript II SK- plasmids were excised from the lambda zap clones by superinfection with Exsist (Stratagene) phage. The library was normalized using method 4 described in Honaldi et al (1996) Genome Research 6: 791-806."

BASE COUNT 153 a 179 c 308 g 117 t
 ORIGIN

Query Match 64.0%; Score 16; DB 170; Length 761;
 Best Local Similarity 100.0%; Pred. No. 52;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 aggtgcatgccagca 23
 |||
 Db 466 AGGTGATGCCAGCA 451

RESULT 7
 BF581354 915 bp mRNA EST 12-DEC-2000
 LOCUS 602100740F1 NCI CGAP_C024 Mus musculus cDNA clone IMAGE:4224225 5',
 DEFINITION mRNA sequence.
 ACCESSION BF581354
 VERSION BF581354.1 GI:11655066
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: c9apbs-r@mail.nih.gov
 Tissue Procurement: Jeffrey E. Green, M.D.
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov

Plate: LLM9813 row: 0 column: 10
High quality sequence stop: 646.
Location/Qualifiers
1. .915

FEATURES

source

/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone_image="4224225"
/clone_lib="NCI_CGAP_Co24"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: colon; Vector: PCMV-SpOxRT; Site: 1: NCI;
Site: 2: Sali; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.6 kb. constructed by Lile
Technologies. Note: this is a NCI_CGAP library."

BASE COUNT

218 a 259 c 230 g 208 t

ORIGIN

Query Match 64.0%; Score 16; DB 150; Length 915;
Best Local Similarity 100.0%; Pred. No. 53;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 accaggtgcatgccca 20
|||||

Db 737 AGCAGTGCATGCCCA 752

RESULT 8
LOCUS BF836559 164 bp mRNA EST 13-JAN-2001
DEFINITION CM2-HT0968-181100-508-e11 HT0968 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF836559
VERSION BF836559.1 GI:12188510
KEYWORDS EST.

SOURCE

human

REFERENCE

AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 164)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,
Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.

Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

JOURNAL

MEDLINE

COMMENT

Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/ILICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=CM2&t2=CM2-HT0968-
181100-508-e11&t3=2000-11-18&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 95
High quality sequence stop: 164.
Location/Qualifiers
1. .164

FEATURES

source

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="HT0968"
/dev_stage="Adult"
/note="Organ: head-neck; Vector: puc18; Site: 1: SmaI;
Site: 2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application

BASE COUNT

No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

46 a 39 c 45 g 34 t

ORIGIN

Query Match

60.0%; Score 15; DB 170; Length 164;

Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gccagcaggtgcat 15
|||||

Db 28 GCCAGCAGGTGCAT 14

RESULT 9

LOCUS

A0851231

DEFINITION

IMAUFV1_lm44h03.x1 Leishmania major FV1 random genomic library
Leishmania major genomic clone IMAUFV1_lm44h03 3', DNA sequence.

ACCESSION

A0851231 GI:6055879
GSS.

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

1 (bases 1 to 226)
Akopyants, N.S., Clifton, S.W., Martin, J., Pape, D., Wylie, T., Li, L.,
Kissinger, J., Roos, D.S., Marra, M., Hillier, L., Chinwalla, A.,
Blistain, A., Schmitt, A., Person, B., Theisling, B., Ritey, E., Ronko,
I., Bennett, J., Cole, R., Underwood, K., Cardenas, M., Gibbons, M.,
Harvey, N., McCann, R., Tsagaris, V., Williams, T., Jackson, Y.,
Bowers, Y., Swaller, T., Waterston, R., Wilson, R. and Beverley, S.M.
A survey of the Leishmania major Friedlin strain V1 genome by
shotgun sequencing: a resource for DNA
Mol. Biochem. Parasitol. 113 (2), 337-340 (2001)
Other-GSS: lm44h03.y1
Contact: Akopyants, NS / Beverley, SM
WashU Leishmania Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewartson.wustl.edu

JOURNAL

COMMENT

Library construction: Natalia S. Akopyants, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
If using this information please cite:
N.S. Akopyants and S.M. Beverley 'A survey of the Leishmania major
Friedlin strain V1 genome by shotgun sequencing' and the Washington
University Genome Sequencing Center for information on obtaining
clone material please contact: Natalia S. Akopyants Ph.D.
(natalia@porcim.wustl.edu) and/or Stephen M. Beverley Ph.D.
(beverley@porcim.wustl.edu)
Seq primer: -40up from Glibco
Class: shotgun

High quality sequence stop: 226.
Location/Qualifiers
1. .226

FEATURES

source

/organism="Leishmania major"
/strain="Friedlin strain V1"
/db_xref="taxon:5664"
/clone_image="IMAUFV1_lm44h03"
/clone_lib="Leishmania major FV1 random genomic library"
/lab_host="TOP10 (Invitrogen)"
/note="Vector: pZero-2 (Invitrogen); Site: 1: EcoRV;
Genomic DNA was isolated from stationary phase cells. For
this library, DNA was sheared to give a tight size
distribution of 1-1.5kb fragments. blunt-ended with T4 DNA
polymerase, dephosphorylated with Shrimp Alkaline
phosphatase and ligated into pZero-2 vector's EcoRV site."

BASE COUNT 40 a 67 c 64 g 55 t

ORIGIN

Query Match 60.0%; Score 15; DB 234; Length 226;
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 cgcagcaggtgcatc 16
 |||
 Db 88 CGCAGCAGTGCATC 102

RESULT 10
 A1410504 266 bp mRNA EST 09-FEB-1999
 LOCUS
 DEFINITION EST38797 Normalized rat heart, Bento Soares Rattus sp. cDNA clone
 RHECX29 3' end, mRNA sequence.
 A1410504
 ACCESSION A1410504.1 GI:4254008
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Rattus sp.
 Rattus sp.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 266)
 Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
 Kerlavage,A.R. and Adams,M.D.
 Rat Genome Project: Generation of a Rat EST (RESE) Catalog & Rat
 Gene Index
 Unpublished (1998)
 JOURNAL
 COMMENT
 Contact: Lee, NH
 The Institute for Genomic Research
 9712, Medical Center Drive, Rockville, MD 20850, USA
 Tel: (301)-838-3529
 Fax: (301)-838-0208
 Email: nhlee@ligr.org
 Seq primer: M13-21.
 Location/Qualifiers
 1..266
 /organism="Rattus sp."
 /db_xref="taxon:10118"
 /clone="RHECX29"
 /clone_lib="Normalized rat heart, Bento Soares"
 /note="Organ: heart; Vector: pT73Pac; Site_1: EcoRI;
 Site_2: NotI"
 BASE COUNT 56 a 61 c 64 g 85 t

ORIGIN

Query Match 60.0%; Score 15; DB 19; Length 266;
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 ggtgcatgcagca 23
 |||
 Db 90 GGTGATCGCAGCA 104

RESULT 11
 AA680915 352 bp mRNA EST 10-DEC-1998
 LOCUS
 DEFINITION LmFRA0307 Leishmania major Amastigote full length cDNA library
 Leishmania major cDNA clone Q83 5', mRNA sequence.
 AA680915
 ACCESSION AA680915.1 GI:3886275
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Leishmania major.
 Leishmania major
 Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
 Leishmania.
 1 (bases 1 to 352)

AUTHORS
 TITLE
 JOURNAL
 COMMENT
 Morrish,A.R., Dyall,S.D., Smith,D.F. and Blackwell,J.M.
 Analysis of Leishmania Major Amastigote Expressed Sequence Tags
 Unpublished (1997)
 On Dec 5, 1997 this sequence version replaced gi:2662920.
 Contact: Blackwell JM
 Cambridge Institute for Medical Research
 Wellcome Trust/MRC Building, Addenbrooke's Hospital, Hills Road,
 Cambridge CB2 2XY, UK
 Tel: 01223 336 143
 Fax: 01223 331 206
 Email: jmb37@cus.cam.ac.uk
 PCR Primers
 FORWARD: GTAAACGACGCGCAGT
 BACKWARD: GGAACAGCTATGACATG
 Seq primer: AATTAACTCCTCCTAAGGC
 High quality sequence stop: 352.
 Location/Qualifiers
 1..352
 /organism="Leishmania major"
 /strain="Friedlin"
 /db_xref="taxon:5664"
 /clone="Q83"
 /clone_lib="Leishmania major Amastigote full length cDNA
 library"
 BASE COUNT 70 a 110 c 86 g 86 t

ORIGIN

Query Match 60.0%; Score 15; DB 10; Length 352;
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 cgcaggtgcatgc 18
 |||
 Db 183 CACGAGTGCATCGC 169

RESULT 12
 AA800788 380 bp mRNA EST 30-APR-1998
 LOCUS
 DEFINITION EST190285 Normalized rat lung, Bento Soares Rattus sp. cDNA clone
 RU0AL45 3' end, mRNA sequence.
 AA800788
 ACCESSION AA800788.1 GI:2863743
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Rattus sp.
 Rattus sp.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 380)
 Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
 Kerlavage,A.R. and Adams,M.D.
 Rat Genome Project: Generation of a Rat EST (RESE) Catalog & Rat
 Gene Index
 Unpublished (1998)
 JOURNAL
 COMMENT
 Contact: Lee, NH
 The Institute for Genomic Research
 9712, Medical Center Drive, Rockville, MD 20850, USA
 Tel: (301)-838-3529
 Fax: (301)-838-0208
 Email: nhlee@ligr.org
 Seq primer: M13-21.
 Location/Qualifiers
 1..380
 /organism="Rattus sp."
 /db_xref="ATCC (inhost):2008206"
 /db_xref="taxon:10118"
 /clone="RU0AL45"
 /clone_lib="Normalized rat lung, Bento Soares"
 /note="Organ: lung; Vector: pT73Pac; Site_1: EcoRI;
 Site_2: NotI"
 BASE COUNT 71 a 100 c 106 g 103 t

ORIGIN

Query Match 60.0%; Score 15; DB 12; Length 380;
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 ggtgcatgcacgca 23
 |||||
 Db 41 GGTGATGCCACGCA 55

RESULT 13
 BE429060/c 396 bp mRNA EST 26-JUL-2000
 LOCUS MTD013.H06F990621 ITEC MTD Durum Wheat Root Library Triticum
 DEFINITION turgidum subsp. durum cDNA clone MTD013.H06, mRNA sequence.
 ACCESSION BE429060
 VERSION BE429060.1 GI:9426903
 KEYWORDS EST.
 SOURCE Triticum turgidum subsp. durum
 ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
 Triticaceae; Triticum.
 REFERENCE 1 (bases 1 to 396)
 Anderson,O.A., Appels,R., Bailey,P., Blake,T., Close,T., Cloutier
 S., Dubcovsky,J., Feuillet,C., Gale,M., Graner,A., Gustafson,P.,
 Hermann,R.G., Holton,T., Jacquemin,J.M., Jia,J., Joudrier,P.,
 Langridge,P., Lazo,G.R., Lin,J.J., McGuire,P., Ogihara,Y.,
 Pecchioni,N., Qualset,C., Schuch,W., Selvaraj,G., Shariflou,M.,
 Sorrells,M., Warburton,M., and Wenzel,G.
 International Triticaceae EST Cooperative (ITEC): Production of
 Expressed Sequence Tags for Species of the Triticaceae
 Unpublished (2000)
 CONTACT: Joudrier P
 INRA, Unite de Biochimie et Biologie Moleculaire des Cereales
 2, place VIALA, 34060 Montpellier cedex 01 FRANCE
 Tel: 33 4 99 61 23 84
 Fax: 33 4 99 61 23 48
 Email: joudrier@enscm.inra.fr
 International Triticaceae EST Cooperative (ITEC)
 http://wheat.pw.usda.gov/genome.
 FEATURES
 location/Qualifiers
 1..396
 /organism="Triticum turgidum subsp. durum"
 /db_xref="taxon:4567"
 /clone="MTD013.H06"
 /clone_lib="ITEC MTD Durum Wheat Root Library"
 /tissue_type="root"
 /dev_stage="3-day-old seedling, water-stressed"
 /note="Vector: pSPORT1; T7 primers used. See pSPORT1
 polylinker site 0.3-2.0 kbp average insert size."

BASE COUNT 96 a 87 c 116 g 97 t

ORIGIN

Query Match 60.0%; Score 15; DB 167; Length 396;
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 gcagtcacgcgca 20
 |||||
 Db 75 GCAGTGACATCGCCA 61

RESULT 14
 AA625125 418 bp mRNA EST 02-MAR-1998
 LOCUS af70a04.r1 Soares.NHMPu_S1 Homo sapiens cDNA clone IMAGE:1047342
 DEFINITION 5', mRNA sequence.
 ACCESSION AA625125

VERSION AA625125.1 GI:2537510
 EST.
 KEYWORDS human.
 SOURCE Homo sapiens
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 418)
 Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
 Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin
 J., Moore,B., Schellenberg,K., Stepcoe,M., Tan,F., Theising,B.,
 White,Y., Wylie,T., Waterston,R. and Wilson,R.
 WashU-NCI human EST Project
 Unpublished (1997)
 CONTACT: Wilson RK
 JOURNAL Washington University School of Medicine
 COMMENT 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert Length: 626 Std Error: 0.00
 Seq primer: -28m13 rev2 Et from Amersham
 High quality sequence stop: 408.
 FEATURES
 location/Qualifiers
 1..418
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1047342"
 /clone_lib="Soares.NHMPu_S1"
 /tissue_type="Pooled human melanocyte, fetal heart, and
 placental uterus"
 /lab_host="DH10B"
 /note="Organ: mixed (see below); Vector: pT7T3D-Pac
 (Pharmacia) with a modified polylinker; Site 1: Not I;
 Site 2: Eco RI; Equal amounts of plasmid DNA from three
 normalized libraries (melanocyte 2NBH, pregnant uterus
 NBHPU, and fetal heart NBH19W) were mixed, and ss circles
 were made in vitro. Following HAP purification, this DNA
 was used as tracer in a subtractive hybridization
 reaction. The driver was PCR-amplified cDNAs from pools of
 5,000 clones made from the same 3 libraries. The pools
 consisted of 1.M.A.G.E. clones 260232-265223,
 340488-345479, and 484488-489479."

BASE COUNT 128 a 93 c 114 g 83 t

ORIGIN

Query Match 60.0%; Score 15; DB 9; Length 418;
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 gcgacagtgatc 16
 |||||
 Db 399 GCACAGAGTGATC 413

RESULT 15
 AUI66733/c 453 bp mRNA EST 23-JAN-2001
 LOCUS AUI66733 Rice callus Oryza sativa subsp. japonica cDNA clone C50066
 DEFINITION AUI66733 Rice callus Oryza sativa subsp. japonica cDNA clone C50066
 , mRNA sequence.
 ACCESSION AUI66733
 VERSION AUI66733.1 GI:12405132
 KEYWORDS EST.
 SOURCE Oryza sativa subsp. japonica.
 ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Euphorbiaceae; Oryza.
 REFERENCE 1 (bases 1 to 453)
 Sasaki,T. and Yamamoto,K.
 TITLE Rice cDNA from callus (2001)

JOURNAL
COMMENT

Unpublished (2001)
Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program
2-1-2 Kannondai, Tsukuba
Ibaraki,
Japan 305
Tel: 0298-38-7441
Fax: 0298-38-7468
Email: tsasaki@abrr.affrc.go.jp
PROJECT = "RGP".
C50066_982.

FEATURES
source

Location/Qualifiers
1. 453
/organism="Oryza sativa subsp. japonica"
/strain="cultivar Nipponbare, sub-species Japonica"
/db_xref="taxon:39947"
/clone_lib="C50066"
/clone_lib="Rice callus"
/note="Vector: Bluescript II SK+; Site_1: SalI; Site_2:
NotI; cDNA prepared from rice callus mRNAs by using
oligo(dT) as a primer and ligating to the SalI-NotI site
of Bluescript II SK+ phagemid."

BASE COUNT
ORIGIN
164 a 83 c 140 g 66 t

Query Match 60.0%; Score 15; DB 108; Length 453;

Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 tgcacgcacgacatt 25
|||||
DB 349 TGCAATCGCACGATT 335

RESULT 16
BE446670 463 bp mRNA EST 25-JUL-2000
LOCUS
WHE1139_F07_L13Z5 Wheat etiolated seedling root normalized cDNA
DEFINITION
library Triticum aestivum cDNA clone WHE1139_F07_L13, mRNA
sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

BE446670
BE446670.1 GI:9446232
EST.
bread wheat.
Triticum aestivum
Spermatophyta: Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
; Triticeae; Triticum.
1 (bases 1 to 463)

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Anderson, O.D., Chao, S., Choi, D.W., Close, T.J., Fenton, R.D., Han
, P.S., Hsia, C.C., Kang, Y., Lazo, G.R., Miller, R., Nguyen, H.T.,
Rausch, C.J., Seaton, C.L., Tong, J.C. and Zhang, D.
The structure and function of the expressed portion of the wheat
genomes. Normalized root cDNA library
Unpublished (2000)
Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific
West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105595818
Email: oanderson@pw.usda.gov
Sequence have been trimmed to remove vector sequence and low
quality sequence with phred score less than 20
Seq primer: Stragatene SK primer.

FEATURES
source

1. 463
/organism="Triticum aestivum"
/cultivar="Chinese Spring"
/db_xref="taxon:4565"
/clone="WHE1139_F07_L13"

BASE COUNT
ORIGIN
119 a 98 c 132 g 114 t

Query Match 60.0%; Score 15; DB 167; Length 463;

Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 gcagtgatgcaccca 20
|||||
DB 111 GCAGTGATGCACCA 97

RESULT 17
AA821080 475 bp mRNA EST 25-NOV-1998
LOCUS
GM09622.5prime GM Drosophila melanogaster ovary Bluescript
DEFINITION
Drosophila melanogaster cDNA clone GM09622.5prime, mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AA821080
AA821080.1 GI:2890948
EST.
fruit fly.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Harvey, D., Hong, L., Evans-Holm, M., Pendleton, J., Su, C., Brokstein
, P., Lewis, S. and Rubin, G.M.
BDGP/HMI Drosophila EST Project
Unpublished (1997)
Contact: Harvey D.
G. M. Rubin-Molecular and Cell Biology
University of California Berkeley
539 LSA, Berkeley, CA 94720-3200, USA
Fax: 510 643 9947
Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu
Plate: 96 row: B column: 10
High quality sequence stop: 439.

FEATURES
source

1. 475
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="GM09622"
/clone_lib="GM Drosophila melanogaster ovary Bluescript"
/sex="female"
/dev_stage="newly eclosed females; germlarium-stage 6"
/lab_host="SOLR"
/note="Organ: ovary; Vector: Bluescript SK; Site_1: EcoRI;
Site_2: XhoI; Constructed using Stragatene ZAP-cDNA
synthesis kit. Oligo dT-primed and directionally cloned at
EcoRI and XhoI in Bluescript SK(+/-)"

```

BASE COUNT      128 a      147 c      121 g      79 t
ORIGIN
Query Match      60.0%; Score 15; DB 12; Length 475;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      8 agcagtcacgcagc 22
        |||
Db      344 AGCGTCATCGCCAGC 358

RESULT 18
Bg101817      475 bp      mRNA      EST      30-JAN-2001
LOCUS      RH122.22.G09.g1.A003 Rhizome2 (RH122) Sorghum prolinguam cDNA, mRNA
DEFINITION      sequence.
ACCESSION      Bg101817
VERSION      Bg101817.1 GI:12616762
KEYWORDS      EST.
SOURCE      Sorghum prolinguam.
ORGANISM      Sorghum prolinguam.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Sorghum.
1 (bases 1 to 475)
Cordonier-Pratt, M.-M., Gingle, A., Paterson, A., Sudman, M. and Pratt
, L.H.
TITLE      An EST database from Sorghum: Sorghum prolinguam rhizomes
JOURNAL      Unpublished (2000)
COMMENT      Contact: Cordonier-Pratt MM
              Department of Botany
              The University of Georgia
              Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
              Tel: 706 542 1860
              Fax: 706 542 1805
              Email: mmpratt@uga.edu
              Seq primer: PolYTm1x
              High quality sequence start: 11
              High quality sequence stop: 408
              POLYA-No.
FEATURES
    source      Location/Qualifiers
                1..475
                /organism="Sorghum prolinguam"
                /db_xref="taxon:132711"
                /clone_lib="Rhizome2 (RH122)"
                /note="Organ: Rhizomes; Vector: pBluescript II from Lambda
                zap II; Site_1: XhoI; Site_2: EcoRI; The library was made
                from poly-A RNA in the cloning vector Lambda ZAP II.
                Clones to be sequenced by mass excision."
BASE COUNT      128 a      107 c      113 g      127 t
ORIGIN
Query Match      60.0%; Score 15; DB 173; Length 475;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      6 gcagtcacgcagc 20
        |||
Db      240 GCAGTCATCGCCCA 254

RESULT 19
AA698088      487 bp      mRNA      EST      27-NOV-1998
LOCUS      HL03757.5prtime HL Drosophila melanogaster head Bluescript
DEFINITION      Drosophila melanogaster cDNA clone HL03757 5prtime, mRNA sequence.
ACCESSION      AA698088
VERSION      AA698088.1 GI:2701017
KEYWORDS      EST.
SOURCE      fruit fly.

```

```

ORGANISM      Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 487)
Harvey, D., Hong, L., Evans-Holm, M., Pendleton, J., Su, C., Brokstein
, P., Lewis, S. and Rubin, G.M.
BDGP/HMMI Drosophila EST Project
Unpublished (1997)
Contact: Harvey, D.
G. M. Rubin-Molecular and Cell Biology
University of California Berkeley
539 USA, Berkeley, CA 94720-3200, USA
Fax: 510 643 9947
Email: http://www.fruitfly.org/EST\_estefruitfly.berkeley.edu
Plate: 37 row: E column: 9
High quality sequence stop: 331.
FEATURES
    source      Location/Qualifiers
                1..487
                /organism="Drosophila melanogaster"
                /db_xref="taxon:7227"
                /clone_lib="HL03757"
                /clone_lib="HL Drosophila melanogaster head Bluescript"
                /sex="male and female"
                /dev_stage="adult"
                /lab_host="SOLR"
                /note="Organ: head-brain & sensory organ; Vector:
                Bluescript SK; Site_1: EcoRI; Site_2: XhoI; Constructed
                using Stratagene ZAP-cDNA Synthesis kit. Oligo dT primed
                and directionally cloned at EcoRI and XhoI in Bluescript
                SK(+/-)"
BASE COUNT      123 a      151 c      128 g      85 t
ORIGIN
Query Match      60.0%; Score 15; DB 10; Length 487;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      5 agcagtcacgcagc 19
        |||
Db      109 AGCAGTCATCGCC 95

RESULT 20
BF916376      499 bp      mRNA      EST      18-JAN-2001
LOCUS      IL3-UT0115-111200-377-F05 UT0115 Homo sapiens cDNA, mRNA sequence.
DEFINITION      BF916376
ACCESSION      BF916376.1 GI:12307834
VERSION      EST.
KEYWORDS      human.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 499)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,
Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare
, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001

```

Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL:
 (http://www.ludwig.org.br/scripts/gethtml2.pl?PI=IL3&L2=IL3-UT0115-111200-377-F05&L3=2000-12-11&L4=1)

Seq primer: puc 18 forward
 High quality sequence stop: 496.

FEATURES

Source

1. 499

Location/Qualifiers

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="UT0115"

/dev_stage="Adult"

/note="Organ: uterus; tumor; Vector: puc18; Site_1: Sma1; Site_2: Sma1; A mini-library was made by cloning products derived from ORESSES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 121 a 121 c 114 g 143 t
 ORIGIN

Query Match 60.0%; Score 15; DB 171; Length 499;
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 6 gccagtcacgcaca 20
 ||||||||||||
 Db 216 GCAGTCATCGCCCA 230

RESULT 21

C98143/c

LOCUS

DEFINITION

C98143 505 bp mRNA EST 29-APR-1999
 C98143 Rice callus Oryza sativa subsp. japonica cDNA clone C0777_7A
 , mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Oryza sativa subsp. japonica.
 Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Ehrhartoideae; Oryzaceae; Oryza.
 1 (bases 1 to 505)

REFERENCE

AUTHORS

TITLE

JOURNAL

Unpublished (1998)
 On Oct 19, 1998 this sequence version replaced g1:454479.
 Contact: Takuji Sasaki
 National Institute of Agrobiological Resources
 Rice Genome Research Program
 2-1-2 Kannondai, Tsukuba
 Ibaraki,
 Japan 305
 Tel: 0298-38-7441
 Fax: 0298-38-7468
 Email: tsasaki@affrc.go.jp
 D8202: Submitted (01-JAN-1994)
 PROJECT = "RGP"

COMMENT

POLYA-No.
 Location/Qualifiers
 1. 505
 /organism="Oryza sativa subsp. japonica"
 /strain="cultivar Nipponbare, sub-species japonica"
 /db_xref="taxon:39947"
 /clone="C0777_7A"
 /clone_lib="Rice callus"
 /note="Vector: pluscript II SK+; Site_1: SalI; Site_2:
 NotI; cDNA prepared from rice callus mRNAs by using
 oligo(dT) as a primer and ligating to the SalI-NotI site

FEATURES
 Source

BASE COUNT 158 a 103 c 170 g 72 t 2 others
 ORIGIN

Query Match 60.0%; Score 15; DB 156; Length 505;
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 11 tgcacgcacagcatt 25
 ||||||||||||
 Db 475 TGCATCGCCAGCATT 461

RESULT 22

B6467085

LOCUS

DEFINITION

B6467085 514 bp mRNA EST 20-MAR-2001
 1A04C06 Bovine Mixed Adipose cDNA library Bos taurus cDNA 5', mRNA
 sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Bos taurus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 Bovidae; Bovinae; Bos.
 1 (bases 1 to 514)

REFERENCE

AUTHORS

TITLE

CDNA's from bovine subcutaneous adipose tissue
 Unpublished (2001)
 Contact: Dr. Stephen Moore
 Beef Genomics Laboratory
 Dept of AFNS, University of Alberta
 410 Agril/Fer, Dept of AFNS, U of A, Edmonton, AB, T6G 2P5, Canada
 Tel: 780 492 0169
 Fax: 780 492 4265

COMMENT

Email: smoores@afns.ualberta.ca
 The sequence best matches gb:BTNADHDUB (B.taurus CI-B14.5b mRNA for
 NADH dehydrogenase (ubiquinone)) in main database at high score of
 908.0 and E-value of 0.0
 PCR Primers
 FORWARD: M13 Forward
 BACKWARD: M13 Reverse
 Seq primer: T3 primer
 High quality sequence stop: 514
 POLYA-No.

FEATURES

Source

Location/Qualifiers
 1. 514
 /organism="Bos taurus"
 /db_xref="taxon:9913"
 /clone_lib="Bovine Mixed Adipose cDNA library"
 /sex="two males and one female mixed"
 /tissue_type="Adipose"
 /cell_type="Adipocyte"
 /dev_stage="young adult"
 /lab_host="XLI-BlueMR/scrain"
 /note="Organ: Subcutaneous, Omental, Mesenteric adipose;
 Vector: Uni-ZAPXR; Site_1: EcoRI; Site_2: Xho I"

BASE COUNT 115 a 112 c 132 g 155 t
 ORIGIN

Query Match 60.0%; Score 15; DB 154; Length 514;
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 11 tgcacgcacagcatt 25
 ||||||||||||
 Db 242 TGCATCGCCAGCATT 256

RESULT 23	AI957283	524 bp	mrna	EST	20-AUG-1999
LOCUS	U18612.x1	Sugano mouse kidney mkia	Mus musculus	cdna clone	
DEFINITION	IMAGE:2158894 3', mRNA sequence.				
ACCESSION	AI957283				
VERSION	AI957283.1	GI:5749992			
KEYWORDS	EST.				
SOURCE	house mouse.				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.				
AUTHORS	1 (bases 1 to 524) Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Smaller,T., Gibbons,M., Page,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,M., Cardenas,M., McCann,R., Waterston,R., and Wilson,R. The WashU-NCI Mouse EST Project 1999 Unpublished (1999) Contract: Marra M/WashU-NCI Mouse EST Project 1999 Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@watson.wustl.edu This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. MGI:1002066				
TITLE	Seq primer: custom primer used				
JOURNAL	High quality sequence stop: 359.				
COMMENT	Location/Qualifiers 1..524 /organism="Mus musculus" /strain="C57BL" /db_xref="taxon:10090" /clone="IMAGE:2158894" /clone_lib="Sugano mouse kidney mkia" /sex="female" /dev_stage="adult" /lab_host="DH10B" /note="Organ: kidney; Vector: pME18S-FL3; Site.1: Draili (CACGTGTG); Site.2: Draili (CACCAATGTG); 1st strand cDNA was primed with an oligo(dT) primer [ATGTGGCCCTTTTCTTTTCTTTT]; double-stranded cDNA was ligated to a Draili adaptor (TGTGGCCCTACTGG), digested and cloned into distinct draili sites of the pME18S-FL3 vector (5' site CACGTGTG, 3' site CACCAATGTG). XhoI should be used to excise the cDNA insert. Size selection was performed to exclude fragments <1.5kb. Library constructed by Dr. Sumio Sugano (University of Tokyo Institute of Medical Science). Custom primers for sequencing: 5' end primer CTTCTGCTCTATAAAGCTGTG and 3' end primer CGACCTGTACACTCGACACA."				
BASE COUNT	151 a 102 c 147 g 124 t				
ORIGIN					
Query Match	60.0%; Score 15; DB 104; Length 524;				
Best Local Similarity	100.0%; Pred. No. 1.9e+02;				
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
QY 11	tgcatcgccacgacatt 25 				
Db 352	TGCATCGCCACGACATT 338				
RESULT 24	BG232022	536 bp	mrna	EST	09-FEB-2001
LOCUS	nat335606.x1	Soares.NPBMG	Homo sapiens	cdna clone	IMAGE:414059 3'
DEFINITION	similar to tr.P79434 P79434 CONNECTING PEPTIDE/TRANSMEMBRANE /CYTOPLASMIC DOMAIN ;, mRNA sequence.				

ACCESSION	BG232022	GI:12727177
VERSION	BG232022.1	
KEYWORDS	EST.	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
TITLE	1 (bases 1 to 536)	
JOURNAL	NCI-CCAP http://www.ncbi.nlm.nih.gov/nciccap .	
COMMENT	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index Unpublished (1997) Contact: Robert Strausberg, Ph.D. Email: cgapbs-r@mail.nih.gov Tissue Procurement: M. Bento Soares, Ph.D. cDNA Library Preparation: M. Bento Soares, Ph.D. and M.Fatima Bonaldo, Ph.D. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CCAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: info@image.lnl.gov Seq primer: -400p from Gibco High quality sequence stop: 412. Location/Qualifiers 1..536 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:4143059" /clone_lib="Soares_NBNC" /tissue_type="Lymphocyte" /lab_host="DH10B (phage-resistant)" /note="Organ: blood; Vector: pT73D-Pac; Site:1: NotI; Site_2: EcoRI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTACCAACTGTAAGTGGAGCGCGCGGCTTTTTTTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library is normalized; constructed in the laboratory of M. Bento Soares (University of Iowa)."	
BASE COUNT	152 a 119 c 133 g 132 t	
ORIGIN		
Query Match	60.0%; Score 15; DB 174; Length 536;	
Best Local Similarity	100.0%; Pred. No. 1.9e+02;	
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
Oy	6 gaagtgcatgcacca 20	
Db		
	293 GCAGGTGCATGCCTA 279	
RESULT 25		
LOCUS	AZ071920 545 bp DNA GSS 31-MAR-2000	
DEFINITION	RPCT-23-395D19.TJB RPCT-23 Mus musculus genomic clone	
ACCESSION	RPCT-23-395D19, DNA sequence.	
VERSION	AZ071920	
KEYWORDS	AZ071920.1 GI:7364817	
SOURCE	GSS.	
ORGANISM	house mouse.	
	Mus musculus	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
REFERENCE	1 (bases 1 to 545)	
AUTHORS	Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akimret B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P. and Fraser,C.M.	
TITLE	Mouse BAC End Sequences from Library RPCT-23	
JOURNAL	Unpublished (1999)	
COMMENT	Other_GSSs: RPCT-23-395D19.TJV	

Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>) or from Resea ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/tldb/bac_ends/mouse/Bac_end_intro.html
Plate: 395 row: D column: 19
Seq primer: SP6
Class: BAC ends.

FEATURES

source

Location/Qualifiers
1..545
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-395D19"
/clone_1lb="RPCI-23"
/sex="Female"
/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBAC3.6; Site: 1; EcoRI; Site: 2; EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBAC3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 137 a 153 c 121 g 134 t
ORIGIN

Query Match 60.0%; Score 15; DB 237; Length 545;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 6 gcaggtgcatgcgca 20
|||||
Db 505 GCAGTGCATGCCCA 491

RESULT 26 553 bp DNA 27-JUL-2000
A2283519/c
LOCUS RPCI-23-125E16.TJ RPCI-23 Mus musculus genomic clone RPCI-23-125E16
DEFINITION
' DNA sequence.
ACCESSION A2283519
VERSION A2283519.1 GI:9525226
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus.

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 553)
Zhao, S., Nieman, W., Feldjlyum, T., Malek, J., Shatsman, S., Aklouet, B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P., and Fraser, C. M.
Mouse BAC End Sequences from Library RPCI-23
Unpublished (1999)
Other_GSSs: RPCI-23-125E16.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong

TITLE
JOURNAL
COMMENT

(pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>) or from Resea ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/tldb/bac_ends/mouse/Bac_end_intro.html
Plate: 125 row: E column: 16
Seq primer: SP6
Class: BAC ends.

FEATURES

source

Location/Qualifiers
1..553
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-125E16"
/clone_1lb="RPCI-23"
/sex="Female"
/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBAC3.6; Site: 1; EcoRI; Site: 2; EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBAC3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 135 a 156 c 126 g 136 t
ORIGIN

Query Match 60.0%; Score 15; DB 241; Length 553;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 6 gcaggtgcatgcgca 20
|||||
Db 525 GCAGTGCATGCCCA 511

RESULT 27

C98142 567 bp mRNA EST 19-OCT-1998
LOCUS C98142 Rice callus Oryza sativa subsp. japonica cDNA clone
DEFINITION C0777_102, mRNA sequence.

ACCESSION C98142
VERSION C98142.1 GI:3760888
KEYWORDS EST.
SOURCE Oryza sativa subsp. japonica.
ORGANISM Oryza sativa subsp. japonica
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

REFERENCE
AUTHORS Ehhardtidae; Oryzae; Oryza. 1 (bases 1 to 567)
Sasaki, T. and Minobe, Y.
Rice cDNA from callus
Unpublished (1994)
Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program
2-1-2 Kannondai, Tsukuba
Ibaraki,
Japan 305
Tel: 0298-38-7441
Fax: 0298-38-7468
Email: tsasak1@dr.afrc.go.jp
PROJECT "RGP".

FEATURES

source

Location/Qualifiers
1..567
/organism="Oryza sativa subsp. japonica"
/strain="cultivar Nipponbare, sub-species japonica"
/db_xref="taxon:39947"
/clone="C0777_102"
/clone_1lb="Rice callus"
/note="Vector: pBluescript II SK+; Site: 1; SalI; Site: 2; NotI; cDNA prepared from rice callus RNAs by using oligo(dT) as a primer and ligating to the SalI-NotI site

BASE COUNT 190 a 105 c 196 g 75 t 1 others
ORIGIN

Query Match 60.0%; Score 15; DB 156; Length 567;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 tgcacgcgcacatc 25
|||||
Db 472 tgcacgcgcacatc 458

RESULT 28
BF489544/c 575 bp mRNA EST 16-APR-2001
LOCUS
DEFINITION AT255594.Sprime AT Drosophila melanogaster adult testes pOTB7
ACCESSION BF489544
VERSION BF489544.1 GI:11572845
KEYWORDS
SOURCE
ORGANISM
fruit fly.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 575)
Stapleton, M., Broksstein, P., Hong, L., Agbayan, A., Baxter, E., Berman
, B., Carlson, J., Champagne, M., Chavez, C., Chew, M., Dorsett, V., Farfan
Liao, G., Fritse, E., George, R., Gonzalez, M., Guarin, H., Harris, N., Li, P.,
Park, S., Paragas, A., Misra, S., Mungall, C. J., Nuno, J., Pacled, J.,
Celinkier, S. and Rubin, G. M.
Berkeley Drosophila Gene Collection Project
unpublished (2000)
CONTACT: Stapleton, M.
BDGP

TITLE
JOURNAL
COMMENT
Lawrence Berkeley National Lab
One Cyclotron Rd, Berkeley, CA 94720, USA
Fax: 510 486 6798
Email: http://www.fruitfly.org/EST_estefruitfly.berkeley.edu
hit genomic sequence AE003597
Plate: AT.255 row: H column: 10
High quality sequence stop: 536.
Location/Qualifiers
1..575
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="AT25594"
/sex="male"
/dev_stage="0-3 day old Ore-R males"
/lab_host="Plates AT.10-AT.120: DHS-alpha. Plates
AT.121-AT.319: DHS-alpha Tona"
/note="Organ: ADULT testes; Vector: pOTB7; Site: 1: RCOBT;
Site 2: XhoI; The mRNA for the testis library was made
from testes and seminal vesicles hand dissected from 0-3
day old Ore-R males. RNA kindly provided by the lab of
Margaret Fuller. Sized fractionated cDNAs were directly
ligated into pOTB7. Plasmid cDNA library."
BASE COUNT 114 a 167 c 162 g 132 t
ORIGIN

Query Match 60.0%; Score 15; DB 149; Length 575;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 cgcagcagcagcagc 16
|||||
Db 520 CGCAGCAGCAGTCATC 506

RESULT 29
AO783596/c 605 bp DNA GSS 03-AUG-1999
LOCUS
DEFINITION HS_3121_A2_F07_77A CTF Approved Human Genomic Sperm Library D Homo
sapiens genomic clone Plate=3121 COL=14 Row=K, DNA sequence.
ACCESSION AO783596
VERSION AO783596.1 GI:5691150
KEYWORDS
SOURCE
ORGANISM
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 605)
Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T.,
Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and
Hood, L.
Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
99380589
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones may be purchased from Research Genetics (inforesgen.com).
BAC end Web Server: <http://www.husc.washington.edu>
Plate: 3121 row: K column: 14
Seq primer: T7
Class: BAC ends
High quality sequence stop: 605.
Location/Qualifiers
1..605
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=3121 COL=14 Row=K"
/clone_lib="CTF Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelBAC11; BAC Clones in
E-Coli DH10B"

JOURNAL
MEDLINE
COMMENT

FEATURES
source

BASE COUNT 187 a 187 c 123 g 92 t 16 others
ORIGIN

Query Match 60.0%; Score 15; DB 233; Length 605;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 gtgcacgcgcacatc 24
|||||
Db 248 GTGCATCGCAGCAT 234

RESULT 30
BG103212 623 bp mRNA EST 30-JAN-2001
LOCUS
DEFINITION RH122_19_E10_g1_A003 Rhizome2 (RH122) Sorghum prolingum cDNA, mRNA
sequence.
ACCESSION BG103212
VERSION BG103212.1 GI:12618045
KEYWORDS
SOURCE
ORGANISM
Sorghum prolingum.
Sorghum prolingum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Sorghum.
1 (bases 1 to 623)
Cordonnier-Pratt, M.-M., Gingle, A., Paterson, A., Sudman, M. and Pratt
L.H.

TITLE
An EST database from Sorghum: Sorghum prolingum rhizomes

JOURNAL Unpublished (2000)
 COMMENT Contact: Cordonnier-Pratt MM
 Department of Botany
 The University of Georgia
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 542 1805
 Email: mmpratt@uga.edu
 Seq primer: PolYTmix
 High quality sequence start: 48
 High quality sequence stop: 560
 POLYA-No.

FEATURES
 source Location/Qualifiers
 1. 623
 /organism="Sorghum propinquum"
 /db_xref="taxon:132711"
 /clone_lib="Rhizome2 (RH12)"
 /note="Organ: Rhizomes; Vector: pBluescript II from lambda
 zap II; Site_1: XhoI; Site_2: EcoRI; The library was made
 from poly-A RNA in the cloning vector lambda zap II.
 Clones to be sequenced were prepared by mass excision."

BASE COUNT 171 a 140 c 149 g 163 t
 ORIGIN

Query Match 60.0%; Score 15; DB 173; Length 623;
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 gcaggtcgcgcga 20
 ||||||||||||
 Db 394 GCAGGTGCATCGCA 408

RESULT 31
 LOCUS BC479445 730 bp mRNA EST 21-MAR-2001
 DEFINITION 602525920f1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:4649525 5',
 mRNA sequence.
 ACCESSION BC479445
 VERSION BC479445.1 GI:13411724
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 730)
 NIH-MGC http://mgs.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgabs-rt@mail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Ling Hong/Rubin Laboratory
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LCM1429 row: h column: 06
 High quality sequence stop: 642.

FEATURES
 source Location/Qualifiers
 1. 730
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4649525"
 /clone_lib="NIH MGC 21"
 /tissue_type="choriocarcinoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: placenta; Vector: pOT7; Site_1: XhoI;
 Site_2: EcoRI; cDNA made by oligo-dT priming.
 directionally cloned into EcoRI/XhoI sites using the
 following 5' adaptor: GGCACGAC(g). Size-selected >500bp

for average insert size 1.8kb. Library constructed by
 Ling Hong in the laboratory of Gerald M. Rubin (University
 of California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies)."
 BASE COUNT 169 a 184 c 193 g 184 t
 ORIGIN

Query Match 60.0%; Score 15; DB 154; Length 730;
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 cgcagcaggtgcac 16
 ||||||||||||
 Db 706 CGCAGCAGGTGCATC 692

RESULT 32
 LOCUS BE194330 761 bp mRNA EST 02-MAR-2001
 DEFINITION HVSMH0085C19f Hordeum vulgare 5-45 DAP spike EST library
 HVCDNA0009 (5 to 45 DAP) Hordeum vulgare cDNA clone HVSMH0085C19f,
 mRNA sequence.
 ACCESSION BE194330
 VERSION BE194330.2 GI:13187270
 KEYWORDS EST.
 SOURCE barley.
 ORGANISM Hordeum vulgare
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
 1 (bases 1 to 761)
 Trilicase; Hordeum.
 Wing, R., Close, T.J., Kleinbols, A., Wise, R., Begum, D., Fritsch, D., Yu
 'T., Anderson, H., Dale, J., Henry, D., Kennodie, S., Palmer, M., Rambo
 'T., Saski, C., Schwartzbeck, J., Simmons, J., Choi, D.W., Main, D. and
 Wood, T.
 Development of a genetically and physically anchored EST resource
 for barley genomics
 Unpublished (2000)
 On Jun 26, 2000 this sequence version replaced gi:8706516.

JOURNAL
 COMMENT On Jun 26, 2000 this sequence version replaced gi:8706516.
 TITLE Contact: Wing RA
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 7288
 Fax: 864 656 4293
 Email: twing@clemson.edu
 Seq primer: AATTAACTCTCACTAAGG
 High quality sequence stop: 647.

FEATURES
 source Location/Qualifiers
 1. 761
 /organism="Hordeum vulgare"
 /cultivar="Morex"
 /db_xref="taxon:4513"
 /clone="HVSMH0085C19f"
 /clone_lib="Hordeum vulgare 5-45 DAP spike EST library
 HVCDNA0009 (5 to 45 DAP)"
 /tissue_type="5-45 DAP Spike"
 /lab_host="SOLR"
 /note="Vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 198 a 150 c 201 g 212 t
 ORIGIN

Query Match 60.0%; Score 15; DB 164; Length 761;
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 gtgcacgcacacat 24
 ||||||||||||
 Db 54 GTGCATCGCACACAT 40

Seq primer: AATTAACTCTACTAAGG
High quality sequence stop: 657.
Location/Qualifiers
1. 824

FEATURES

/organism="Hordene vulgare"
/cultivar="Moxex"
/db_xref="taxon:4513"
/clone="HVSMEC0004A20f"
/clone_lib="Hordene vulgare seedling shoot EST library
HVCN0003 (Etiolated and unstressed)"
/tissue_type="Seedling shoot"
/lab_host="TUC121"
/note="Vector: lambdaZAP, Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 247 a 177 c 253 g 147 t
ORIGIN

Query Match 60.0%; Score 15; DB 152; Length 824;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ggcagcaggtgcat 15
|||||
DB 801 GCCAGCAGGTGCAT 815

RESULT 36

AJ282984 843 bp mRNA EST 30-JUN-2000
LOCUS
DEFINITION 4A3A-P2G5-R Anopheles gambiae immune competent 4A3A Anopheles
gambiae CDNA clone 4A3A-P2G5, mRNA sequence.

ACCESSION AJ282984
VERSION AJ282984.1 GI:6930863
KEYWORDS
SOURCE EST.

ORGANISM

African malaria mosquito.
Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae
; Anopheles.
1 (bases 1 to 843)

REFERENCE Dimopoulos, G., Casavant, T. L., Chang, S., Scheetz, T., Roberts, C.,
Donohue, M., Schultz, J., Benes, V., Bork, P., Ansoorge, W., Soares, M. B.
and Kafatos, F. C.
Anopheles gambiae pilot gene discovery project: identification of
mosquito innate immunity genes from expressed sequence tags
generated from immune-competent cell lines
Proc. Natl. Acad. Sci. U.S.A. 97 (12), 6619-6624 (2000)

TITLE

CONTACT: Dimopoulos G
Fotis C. Kafatos Laboratory
European Molecular Biology Laboratory
Meyerhofstrasse 1, 69117 Heidelberg, Germany.
Location/Qualifiers
1. 843

COMMENT

/organism="Anopheles gambiae"
/strain="4A r/r"
/db_xref="taxon:7165"
/clone="4A3A-P2G5"
/clone_lib="Anopheles gambiae immune competent 4A3A"
/cell_line="Immune competent 4A3A"
/lab_host="E. coli DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site_1: EcoRI; Site_2: NotI; sequenced from
forward priming site which reads from the 3' end of the
CDNA. The 4A3A is a directionally cloned and normalized
CDNA library that was constructed from the 4A3A cell line
oligo-T primed cDNA according to: Bonaldi, Lennon & Soares
(1996) : Normalization and Subtraction: Two approaches to
facilitate Gene Discovery, Genome Research 6, 791-806."

FEATURES

BASE COUNT 224 a 227 c 231 g 160 t 1 others

ORIGIN

BASE COUNT 224 a 227 c 231 g 160 t 1 others

Query Match 60.0%; Score 15; DB 104; Length 843;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 tgcacgcagcagcat 25
|||||
DB 512 TGCATGCCAGCAT 526

RESULT 37

BE731371 886 bp mRNA EST 15-SEP-2000
LOCUS
DEFINITION 601565029f1 NIH_MGC_21 Homo sapiens CDNA clone IMAGE:3840078 5',
mRNA sequence.

ACCESSION BE731371
VERSION BE731371.1 GI:10145363
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens

REFERENCE

1 (bases 1 to 886)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov

COMMENT

Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
plate: LLCM529 row: 1 column: 07
High quality sequence start: 3
High quality sequence stop: 776.
Location/Qualifiers
1. 886

FEATURES

source

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3840078"
/clone_lib="NIH_MGC_21"
/tissue_type="choriocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: placenta; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; CDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(c). Size-selected
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 151 a 279 c 271 g 185 t
ORIGIN

Query Match 60.0%; Score 15; DB 139; Length 886;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ggcagcaggtgcat 15
|||||
DB 785 GCCAGCAGGTGCAT 799

RESULT 38

BF684687 899 bp mRNA EST 22-DEC-2000
LOCUS
DEFINITION 602141056f1 NIH_MGC_46 Homo sapiens CDNA clone IMAGE:4302204 5',
mRNA sequence.

ACCESSION BF684687
VERSION BF684687.1 GI:11970095
KEYWORDS EST.

SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 899)
AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapsof@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Invitrogen, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Plate: L16M163 row: P column: 13
High quality sequence stop: 584.
Location/Qualifiers
1. 899
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:4302204"
/clone_lib="NIH-MGC_46"
/tissue_type="leiomyosarcoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: uterus; Vector: pGB7; Site:1: XhoI; Site:2: EcoRI. CDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAGCAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-CDNA synthesis kit (stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH-MGC library."

BASE COUNT 171 a 249 c 341 g 138 t
ORIGIN

Query Match 60.0%; Score 15; DB 168; Length 899;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gccagcagcagtcac 15
|||||
Db 830 GCCAGCAGGTCAT 844

RESULT 39
BE709927 116 bp mRNA EST 12-SEP-2000
LOCUS BE709927
DEFINITION IL3-H70619-280600-192-H06 H70619 Homo sapiens CDNA, mRNA sequence.
ACCESSION BE709927
VERSION BE709927.1 GI:10098296
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 116)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome project. This entry can be seen in the following URL
(<http://www.ludwig.org.br/scripts/gethtml2.pl?l=6t2-1L3-H70619-280600-192-H06&t3=2000-06-28&t4=1>)
Seq primer: puc 18 forward
High quality sequence start: 57
High quality sequence stop: 115.
Location/Qualifiers
1. 116
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="H70619"
/dev_stage="Adult"
/note="Organ: head_neck; Vector: puc18; Site:1: SmaI; Site:2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and CDNA amplification were performed under low stringency conditions."

BASE COUNT 27 a 34 c 32 g 23 t
ORIGIN

Query Match 56.0%; Score 14; DB 139; Length 116;
Best Local Similarity 100.0%; Pred. No. 6.4e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gccagcagcagtcac 14
|||||
Db 51 GCCAGCAGGTCGA 64

RESULT 40
BF876173 119 bp mRNA EST 17-JAN-2001
LOCUS BF876173
DEFINITION CM4-ET0097-111100-426-H09 ET0097 Homo sapiens CDNA, mRNA sequence.
ACCESSION BF876173
VERSION BF876173.1 GI:12266303
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 119)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome project. This entry can be seen in the following URL
(<http://www.ludwig.org.br/scripts/gethtml2.pl?l=CM4&t2=CM4-ET0097-111100-426-H09&t3=2000-11-11&t4=1>)

Seq primer: puc 18 forward
High quality sequence start: 15
High quality sequence stop: 118.
Location/Qualifiers

FEATURES

source

1. 119
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="ET0097"
/dev_stage="Adult"
/note="Organ: lung; tumor; Vector: puc18; Site: 1; SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from OPRESSES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT 29 a 24 c 35 g 31 t
ORIGIN

Query Match 56.0%; Score 14; DB 170; Length 119;
Best Local Similarity 100.0%; Pred. No. 6.4e+02;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 cagcaggtgcacg 17
|||||

Db 7 CAGCAGGTGCATCG 20

RESULT 41

A259503/c

LOCUS

1M0407F22F Mouse 10kb plasmid UUCG1M library Mus musculus genomic
clone UUCG1M0407F22 F, DNA sequence.

ACCESSION

A259503

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 178)
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,
M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.
and Wright, D., Weiss, R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert length: 10000 Std Error: 0.00
Plate: 0407 row: F column: 22
Seq primer: CGTGTAAACGACGCCACG
Class: plasmid
High quality sequence stop: 178.
Location/Qualifiers
1. 178
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUCG1M0407F22"
/clone_lib="Mouse 10kb plasmid UUCG1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PMD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson

Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptored DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pMD42 (91147321419b1AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptored mouse DNA was annealed to
adaptored vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

BASE COUNT 33 a 46 c 47 g 52 t
ORIGIN

Query Match 56.0%; Score 14; DB 246; Length 178;
Best Local Similarity 100.0%; Pred. No. 6.5e+02;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 gcagcaggtgcacg 16
|||||

Db 115 GCAGCAGGTGCATCG 102

RESULT 42

BF017303/c

LOCUS

BF017303.x1 Soares_mammary_gland_NMLMG Mus musculus cDNA clone
IMAGE:3467679 3' similar to TR:070278 070278 MULTIPLE ENDOCRINE
NEOPLASIA TYPE 1 CANDIDATE PROTEIN NUMBER 18. ;, mRNA sequence.

ACCESSION

BF017303

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 196)
NCI-CCG http://www.ncbi.nlm.nih.gov/ccgacp.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: c9aps-r@mail.nih.gov
This clone is available royalty-free through INMIL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:1387039
Trace considered overall poor quality
High quality sequence stop: 1.
Location/Qualifiers
1. 196
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:3467679"
/clone_lib="Soares_mammary_gland_NMLMG"
/sex="female (lactating)"
/tissue_type="mammary gland"
/lab_host="DH10B"
/note="Vector: pTR73D-Pac (Pharmacia) with a modified
polylinker; 1st strand cDNA was prepared from mammary
gland tissue from a lactating female, and was then primed
with a Not I - Oligo(dT) primer. Double-stranded cDNA was
ligated to Eco RI adaptors (Pharmacia), digested with Not
I and cloned into the Not I and Eco RI sites of the
modified pTR73 vector. Library is normalized. Library
was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 24 a 53 c 58 g 61 t

ORIGIN

Query Match 56.0%; Score 14; DB 143; Length 196;
 Best Local Similarity 100.0%; Pred. No. 6.5e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 agcaggtgcatgc 18
 |||||
 Db 186 AGCAGTGCATCGC 173

RESULT 43

LOCUS AT1714341 206 bp mRNA EST 08-JUN-1999
 DEFINITION UI-R-AF1-aay-e-12-0-UI.s1 UI-R-AF1 Rattus norvegicus cDNA clone
 UI-R-AF1-aay-e-12-0-UI 3', mRNA sequence.
 AT1714341
 ACCESSION AT1714341.1 GI:5018141
 VERSION EST
 KEYWORDS Norway rat.
 SOURCE Rattus norvegicus
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE 1 (bases 1 to 206)
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene
 JOURNAL discovery
 MEDLINE Genome Res. 6 (9), 791-806 (1996)
 COMMENT Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msoares@blue.weeg.iowa.edu

The sequence contained an oligo-dT track that was present in the
 strand cDNA and therefore this may represent a bonafide poly A
 tail. The sequence tag present in the cDNA between the NotI site
 and the oligo-dT track served to verify it as a clone from the
 M.B. Soares Lab Clone library cDNA library preparation:
 through Research Genetics (www.resgen.com) will be available
 Seq primer: M13 Forward
 POLVA=Yes.

FEATURES

Source

Location/Qualifiers
 1..206
 /organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone="UI-R-AF1-aay-e-12-0-UI"
 /clone_lib="UI-R-AF1"
 /dev_stage="adult"
 /lab_host="DHI0B (Life Technologies)"
 /note="Vector: p773D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-AF1
 library is a normalized library constructed from 15 dpc
 rat atrioventricular (AV) canal. The tag is a string of 5
 nucleotides present between the Not I site and the
 oligo-dT track. The library was constructed as described
 by Bonaldo, Lennon and Soares, Genome Research 6: 791-806
 , 1996. Tissue provided by Jim Lin, Department of Biology,
 University of Iowa.
 TAG_LIB=UI-R-AF1
 TAG_TISSUE=AV canal at 15 dpc
 TAG_SEQ=CAAGG"
 51 a 58 c 41 g 56 t

FEATURES

Source

Query Match 56.0%; Score 14; DB 24; Length 206;
 Best Local Similarity 100.0%; Pred. No. 6.5e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 ggcagcaggtgcatc 16
 |||||
 Db 149 GCAGCAGTGCATC 136

RESULT 44

LOCUS BB563629 219 bp mRNA EST 29-NOV-2000
 DEFINITION BB563629 RIKEN full-length enriched, 18 days embryo Mus musculus
 cDNA clone 1110002M02 5', mRNA sequence.
 BB563629
 ACCESSION BB563629.1 GI:11454521
 VERSION EST
 KEYWORDS house mouse.
 SOURCE Mus musculus
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 219)

REFERENCE 1 (bases 1 to 219)
 AUTHORS Aizawa,K., Akahira,S., Akimura,T., Arai,A., Arakawa,T., Carninci,P.,
 Hanagaki,T., Hayatsu,N., Hiraoka,T., Hirozane,T., Hodayama,Y.,
 Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kawai,J., Kojima,Y., Kono
 ,H., Kusakabe,M., Matsuyama,T., Miyazaki,A., Nakamura,M., Nishi,K.,
 Nomura,K., Numazaki,R., Okazaki,Y., Okido,T., Owa,C., Sakai,C.,
 Sakai,K., Sasaki,D., Sato,K., Shibata,K., Shibata,Y., Shingawa,A.,
 Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Tanaka
 ,T., Toyota,T., Warahiki,A., Yamamura,T., Yasunishi,A., Yoshida,K.,
 Yoshiki,A., Muramatsu,M. and Hayashizaki,Y.
 RIKEN Mouse ESTs (Aizawa,K. et al. 2000)
 Unpublished (2000)
 CONTACT: Yoshinhide Hayashizaki
 Genome Exploration Research Group, Life Science Tsukuba Center,
 The Institute of Physical and Chemical Research (RIKEN), Genomic
 Sciences Center
 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
 Tel: +81-298-36-9013
 Fax: +81-298-36-9098
 Email: genome-res@rtc.riken.go.jp,
 URL: http://genome.rtc.riken.go.jp/
 Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoaka,S., Sasaki
 ,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
 Thermostabilization and thermoactivation of thermostable enzymes by
 trehalose and its application for the synthesis of full length
 cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
 Itoh,M., Katsunari,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
 Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
 ,Y. and Hayashizaki,Y.
 Automated filtration-based high-throughput plasmid preparation
 system. Genome Res. 9 (5), 463-470 (1999)
 Carninci,P. and Hayashizaki,Y.
 High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
 19-44 (1999)
 Please visit our web site (http://genome.rtc.riken.go.jp) for
 further details.

Location/Qualifiers
 1..219
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="1110002M02"
 /clone_lib="RIKEN full-length enriched, 18 days embryo"
 /dev_stage="18 days embryo"
 /lab_host="SOLR"
 /note="Site_1: XhoI; Site_2: SstI; cDNA library was
 prepared and sequenced in Mouse Genome Encyclopedia
 Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in
 RIKEN. Division of Experimental Animal Research in Riken
 contributed to prepare mouse tissues. 1st strand cDNA was
 primed with a primer [5'

Search completed: October 9, 2001, 15:15:31
Job time: 13653 sec

GAGAGAGAGCGCGCGCACTGAGATTTTCTTTTCTTTTNN 3'}, cDNA was prepared by using trehalase thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence 15'

BASE COUNT

46 a 76 c 46 g 51 t

ORIGIN

Query Match

Best Local Similarity 56.0%; Score 14; DB 161; Length 219;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcgcagcaggtgca 14

Db 57 GCGCAGCAGGTGCA 70

RESULT 45

AM326327

LOCUS

DEFINITION

AM326327

AM326327

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

1 (bases 1 to 226)

Fahrenkrug, S.C., Freking, B.A., Rohrer, G.A., Smith, T.P.L., Casas, E., Stone, R.T., Heaton, M.P., Grose, W.M., Bennett, G.A., Laegreid, W.W. and Keele, J.W.

Design and use of two pooled tissue normalized cDNA libraries for EST discovery in swine

Unpublished (2000)

Contact: Smith TPL

USDA, ARS, US Meat Animal Research Center

PO Box 166, Clay Center, NE 68933-0166, USA

Tel: 402 762 4366

Fax: 402 762 4390

Email: smithemail.marc.usda.gov

Single pass sequencing. Bases called and trimmed with phred

v0.980904.e. Vector identified by cross-match with the -minscore 20

and -mismatch 12 options.

PCR Primers

FORWARD: AGGAACAGCTATGACCAT

BACKWARD: GTTTCACAGTCAGACG

Plate: 10 Row: K Column: 8

Seq primer: ATTAGTGACACTATAG.

Location/Qualifiers

1..226

/organism="Sus scrofa"

/db_xref="taxon:9823"

/clone_lib="MARC 2P1G"

/tissue_type="pooled"

/lab_host="DH10B"

/note="Vector: PCMW SPORT6; Site_1: XbaI; Site_2: XhoI;

library made from pooled tissue from testis, ovary,

endometrium, hypothalamus, pituitary, and placenta."

BASE COUNT

40 a 71 c 75 g 40 t

ORIGIN

Query Match

Best Local Similarity 56.0%; Score 14; DB 114; Length 226;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcgcagcaggtgca 14

Db 173 GCGCAGCAGGTGCA 186

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 15:45:48 ; Search time 1670.83 Seconds
(without alignments)
231.438 Million cell updates/sec

Title: US-09-396-196f-6

Perfect score: 25
Sequence: 1 gcaggtgcatgcgccagattcgat 25

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1344157 seqs, 773874588 residues

Word size : 9

Total number of hits satisfying chosen parameters: 207944

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl:**

1: gb_da1:*
2: gb_da2:*
3: gb_da3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rud:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rud:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*

44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_v1:*
53: gp_sts1:*
54: gp_sts2:*
55: gp_sts3:*
56: gp_sy:*
57: gp_un:*
58: gp_v11:*
59: gp_v12:*
60: gp_htg1:*
61: gp_htg2:*
62: gp_htg3:*
63: gp_htg4:*
64: gp_htg5:*
65: gp_htg6:*
66: gp_htg7:*
67: gp_htg8:*
68: gp_htg9:*
69: gp_htg10:*
70: gp_htg11:*
71: gp_htg12:*
72: gp_htg13:*
73: gp_htg14:*
74: gp_htg15:*
75: gp_htg16:*
76: gp_htg17:*
77: gp_htg18:*
78: gp_htg19:*
79: gp_htg20:*
80: gp_htg21:*
81: gp_htg22:*
82: gp_htg23:*
83: gp_htg24:*
84: gp_htg25:*
85: gp_pr1:*
86: gp_pr2:*
87: gp_pr3:*
88: gp_pr4:*
89: gp_pr5:*
90: gp_pr6:*
91: gp_pr7:*
92: gp_pr8:*
93: gp_pr9:*
94: gp_r01:*
95: gp_r02:*
96: gp_in4:*
97: gp_pr10:*
98: em_da3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	100.0	1041	9 AR029499	AR029499 Sequence
2	25	100.0	1041	9 AR034916	AR034916 Sequence
3	25	100.0	1084	9 A11530	A11530 B10b gene O
4	25	100.0	1121	10 E00893	E00893 Genomic DNA
5	25	100.0	5793	2 ECOB10	J04423 E.coli 7,8-
6	25	100.0	5872	9 A38246	A38246 Sequence 1
7	25	100.0	5872	9 A38251	A38251 Sequence 6
8	25	100.0	5872	9 A93674	A93674 Sequence 1

```

9      100.0  5872  9  A93679
10     100.0  5872  9  AR101809
11     100.0  5872  9  AR101810
12     100.0  11022 1  AE000180
13     100.0  13501 1  AE000528
14     100.0  297816 2  AP002553
15     100.0  3603 3  PSEHRIA
16     72.0  2923 3  MAV250020
17     72.0  5526 2  AF250776
18     72.0  479 54  G01446
19     68.0  53067 65  AC019940
20     68.0  153870 63  AC013431
21     68.0  183937 71  AC037435
22     68.0  327446 4  AE003500
23     64.0  1186 2  AF146614
24     64.0  4674 89  AK024397
25     64.0  11910 56  AF206717
26     64.0  49587 6  DMBH48C10
27     64.0  60573 65  AC018267
28     64.0  82289 13  ATAC009895
29     64.0  117968 92  HS272E8
30     64.0  140825 13  AP002817
31     64.0  146081 13  AP001366
32     64.0  154305 71  AC034114
33     64.0  177380 85  AC002457
34     64.0  180700 76  AC084136
35     64.0  183439 4  AC007809
36     64.0  232744 5  AE003705
37     64.0  237119 65  AC017740
38     64.0  289090 4  AE003424
39     60.0  1389 14  NTPPIG
40     60.0  1713 14  NTEPCL1
41     60.0  1907 93  HSSVNTRO5
42     60.0  2616 85  AB048915
43     60.0  3369 13  AF274670
44     60.0  3369 13  AF274671
45     60.0  3677 3  PSEIAAHM

```

ALIGNMENTS

```

RESULT 1
LOCUS   AR029499 1041 bp DNA
DEFINITION Sequence 7 from patent US 5859335.
ACCESSION AR029499
VERSION   AR029499.1 GI:5941472
KEYWORDS
SOURCE    Unknown.
ORGANISM  Unknown.
REFERENCE 1 (bases 1 to 1041)
AUTHORS   Patton,D.Andrew.
TITLE     Enhanced biotin biosynthesis in plant tissue
JOURNAL   Patent: US 5859335-A 7 12-JAN-1999;
FEATURES
     source             1..1041
                        /organism="unknown"
BASE COUNT 262 a 273 c 305 g 201 t
ORIGIN

```

```

Query Match
Best Local Similarity 100.0%; Score 25; DB 9; Length 1041;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 gcaagtgatcgccagcattcgat 25
    ||||||||||||||||||||
Db 84 GCAGGTGCATCGCCAGCATTTTCAT 108

```

```

RESULT 2

```

```

AR034916
LOCUS   AR034916 1041 bp DNA
DEFINITION Sequence 7 from patent US 5869719.
ACCESSION AR034916
VERSION   AR034916.1 GI:5950521
KEYWORDS
SOURCE    Unknown.
ORGANISM  Unknown.
REFERENCE 1 (bases 1 to 1041)
AUTHORS   Patton,D.A.
TITLE     Transgenic plants having increased biotin content
JOURNAL   Patent: US 5869719-A 7 09-FEB-1999;
FEATURES
     source             1..1041
                        /organism="unknown"
BASE COUNT 262 a 273 c 305 g 201 t
ORIGIN

```

```

Query Match
Best Local Similarity 100.0%; Score 25; DB 9; Length 1041;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 gcaagtgatcgccagcattcgat 25
    ||||||||||||||||||||
Db 84 GCAGGTGCATCGCCAGCATTTTCAT 108

```

```

RESULT 3
LOCUS   A11530 1084 bp DNA
DEFINITION Biob gene of E.coli with primers.
ACCESSION A11530
VERSION   A11530.1 GI:490218
KEYWORDS
SOURCE    Escherichia coli.
ORGANISM  Escherichia coli.
REFERENCE 1 (bases 1 to 1084)
AUTHORS   Patent: GB 2216530-A 16 11-OCT-1989;
JOURNAL   Location/Qualifiers
FEATURES
     source             1..1084
                        /organism="Escherichia coli"
                        /db_xref="taxon:562"
                        /gene="BIOB"
                        /gene="BIOB"
                        /gene="BIOB"
                        /codon_start=1
                        /transl_table=11
                        /protein_id="CAA00965.1"
                        /db_xref="GI:490219"
                        /db_xref="SWISS-PROT:P12996"
                        /translation="MAHRRKMTLSQVTRLEFKPLDILDFEAQVYHROHPPQOVOST
                        LSIKTAGDEOCKYCROSKRYTGILEARLLEVEOVLESARKARAGSTRPCMAAW
                        KNPHERDMPLLEDMOVGVKAMGLEACMTIGTISEQOARLANAGLDYVNHNDTSPER
                        YGNTITTRKYQERLDITLERVDAGIKVCSGGIVIGETVKDRAGLILQIANLPTPES
                        VPIMLVKKVGTPLADNDVDVDAFDFRTIAVARIMMPTSYVRLSAGREDMNROTQAMC
                        FMAGANSIFYGCKLITTPNPERDKDLQIFRKILGNPQGTAVIAGNEDQQRLEQALMT
                        PTDDEVYNAAL"

```

```

REFERENCE 1 (bases 1 to 1084)
AUTHORS   Patent: GB 2216530-A 16 11-OCT-1989;
JOURNAL   Location/Qualifiers
FEATURES
     source             1..1084
                        /organism="Escherichia coli"
                        /db_xref="taxon:562"
                        /gene="BIOB"
                        /gene="BIOB"
                        /gene="BIOB"
                        /codon_start=1
                        /transl_table=11
                        /protein_id="CAA00965.1"
                        /db_xref="GI:490219"
                        /db_xref="SWISS-PROT:P12996"
                        /translation="MAHRRKMTLSQVTRLEFKPLDILDFEAQVYHROHPPQOVOST
                        LSIKTAGDEOCKYCROSKRYTGILEARLLEVEOVLESARKARAGSTRPCMAAW
                        KNPHERDMPLLEDMOVGVKAMGLEACMTIGTISEQOARLANAGLDYVNHNDTSPER
                        YGNTITTRKYQERLDITLERVDAGIKVCSGGIVIGETVKDRAGLILQIANLPTPES
                        VPIMLVKKVGTPLADNDVDVDAFDFRTIAVARIMMPTSYVRLSAGREDMNROTQAMC
                        FMAGANSIFYGCKLITTPNPERDKDLQIFRKILGNPQGTAVIAGNEDQQRLEQALMT
                        PTDDEVYNAAL"

```

```

BASE COUNT 271 a 286 c 318 g 209 t
ORIGIN

```

```

Query Match
Best Local Similarity 100.0%; Score 25; DB 9; Length 1084;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 gcaagtgatcgccagcattcgat 25
    ||||||||||||||||||||

```

Db 107 GCAGTGCATCGCCAGCATTTGCAT 131

RESULT 4

LOCUS E00893 1121 bp DNA PAT 29-SEP-1997

DEFINITION Genomic DNA encoding biotin Synthetase.

ACCESSION E00893

VERSION E00893.1 GI:2169154

KEYWORDS JP 1986149091-A/1.

SOURCE *Escherichia coli*.

ORGANISM *Escherichia coli*

Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; *Escherichia*.

REFERENCE 1 (bases 1 to 1121)

AUTHORS Hirono, Y., Kojima, T. and Kimura, H.

TITLE DUPLICATION OF BIOTIN SYNTHASE, BACTERIUM CONTAINING SAME AND PRODUCTION OF BIOTIN

JOURNAL Patent: JP 1986149091-A 1 07-JUL-1986;

COMMENT NIPPON SODA CO LTD

OS *Escherichia coli*

PN JP 1986149091-A/1

PD 07-JUL-1986

PF 24-DEC-1984 JP 1984272605

PI HIRONO YOSHIHIKO, KOJIMA TAKAKAZU, KIMURA HITOSHI PC

C12N15/00, C12N1/20, C12P13/18, (C12N1/20, C12R1:19), (C12P13/18, PC

C12R1:19);

CC strandedness: Double;

CC topology: Linear;

CC hypothetical: No;

CC anti-sense: No;

CC *source: strain-*Escherichia coli* N5101;

CC feature is identified by experimental;

FH key Location/Qualifiers

FT CDS 42..1079

FT /product="biotin synthetase".

FEATURES

source Location/Qualifiers

1..1121

/organism="Escherichia coli"

/db_xref="taxon:562"

BASE COUNT 289 a 296 c 325 g 211 t

ORIGIN

Query Match 100.0%; Score 25; DB 10; Length 1121;

Best Local Similarity 100.0%; Pred. No. 0.00037;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 gcagtgcatcgccagcatttcgat 25

111111111111111111111111

Db 125 GCAGTGCATCGCCAGCATTTGCAT 149

RESULT 5

LOCUS E00893 5793 bp DNA BCT 28-FEB-1994

DEFINITION *E. coli* 7,8-diamino-pelargonic acid (bioa), biotin synthetase (bioB), 7-keto-8-amino-pelargonic acid synthetase (bioF), bioc protein, and dehydrobiotin synthetase (biobD), complete cds.

ACCESSION J04423

VERSION J04423.1 GI:145422

KEYWORDS 7,8-diamino-pelargonic acid aminotransferase; 7-keto-8-amino-pelargonic acid synthetase; bioa gene; biob gene; bioc gene; biob gene; biof gene; biotin synthetase; dehydrobiotin synthetase.

SOURCE *Escherichia coli* (strain K-12) DNA.

ORGANISM *Escherichia coli*

Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; *Escherichia*.

REFERENCE 1 (bases 1 to 5793)

AUTHORS Otsuka, A.J., Buoncristiani, M.R., Howard, P.K., Flamm, J. and Johnson, O.

TITLE The *Escherichia coli* biotin biosynthetic enzyme sequences

JOURNAL U. Biol. Chem. 263, 19577-19585 (1988)

MEDLINE 89066784

COMMENT Draft entry and computer-readable sequence [1] kindly submitted by A.Otsuka, 09-NOV-1988.

FEATURES

source Location/Qualifiers

1..5793

/organism="Escherichia coli"

/strain="K-12"

/db_xref="taxon:562"

/complement(98..574)

/note="ORF 1"

/codon_start=1

/transl_table=1

/protein_id="AAA23513.1"

/db_xref="GI:45168"

/translation="MKLISNDLRDGDLPKRRHFWNGMGVDGDNISPHLAMPDVPAQTK SEVVCYDDADPTGSGMHWVYVNLPAQTRVLPQGGSLVAMPDGVLOTPTDEKTKG YDGAAPPKGETHRYLFTYHALDIERIDVDEGASAAVGFNVHFSLASAITAMS"

/complement(633..1925)

/gene="bioa"

/complement(633..1925)

/gene="biob"

/complement(633..1925)

/gene="bioc"

/complement(633..1925)

/transl_table=11

/product="7,8-diamino-pelargonic acid aminotransferase"

/protein_id="AAA23514.1"

/db_xref="GI:457106"

/translation="MTTDLAPDQRIHTPTSMSPPLVPYVVSNECCETILSPGR LVDGMSWMAAIGHVNHQNLNAMSQIDAMSHVMFGITPALELCKRLVMSGRN ALECVFLADSGVAVVAMKMAIQWQAGEARORPLTFRNQYHDDTFCAMSVCPDN SMHSLKGYLPELFAFAPDSRMDGEMDERGAPRLMAARHETIAVITIEPIVGA GGRMYNHPMKIRIKICDREGILADELATGFGFTGKLPACAEANAALADLIGKA LRGCTVLTSLNTITTPREVAETISNGEACGEMHGPTEGNPLVMAALQKRFVQGWIR QQVADLEVQRLAPARDAENYADVRYGALGVYETTHPVMAALQKRFVQGWIR PFQGLIYLPPTIILPQQLRLTAANRAVQDETFQC"

2012..3052

/gene="biob"

2012..3052

/gene="biob"

/gene="biob"

/codon_start=1

/transl_table=11

/product="biotin synthetase"

/protein_id="AAA23515.1"

/db_xref="GI:145425"

/translation="MAHRPMTLSQVTELEFEPDLDLFEAOOVHRHEDPQVQVST LSTKTAGCEDCKYCPQTSRYTGLEARIMEVQVLESARAKAAGSTRQMGAM KNPERMDPTLEOVGVKAMGLEACHTGCTLSESAQRLANAGLDYVNHNDTSPF YGNITTRTYOERLDLLEKVRDAGIRVCSGTVGLDEYKADRAGLLQANLPTPES VPINMLVYKRGFTPLANDVDVADPFRITIAVARITMPTSVRLSAGREOMNEQTMG FMAGANSIFYGCKLITTPNEDEKDLQLEFRKGLINPQQAVALAGDNEQOORLEQALMT PPTDEVYNAAL"

3049..4203

/gene="biob"

3049..4203

/gene="biob"

/gene="biob"

/codon_start=1

/transl_table=11

/product="7-keto-8-amino pelargonic acid synthetase"

/protein_id="AAA23516.1"

/db_xref="GI:145426"

/translation="MSWQERTINAALDARRADALRRRYPVAGCGRMLVADPOYLNF SSNDYLSHHPQIIRAMOGAGDQFGIGSGSGVSGSVVHOLELEELAEMLGYSRA LLTSGFANADVAIAAMAKEDRIADRLSHASILEAASLSPQLARFPAHNDVTHLAR LLSAPCGOQWVYVEGVESMDGSAPLAEIQOYTOOHNGLAMVDAGTGVIGEORG SCHLOKRPBELIVYTPKGGVSGAVALCSITVADYLOFARHLIYSTSPAPAOAL RASLAVIRSDGDAKREKLAALITRPAQVODLPFTLADSCSAIOLPIVDSRAQLL AEKLRQCGCWYTAIRPPTVAGTARLRLTLAAHEMODIRLLEVLHGNG"

4190..4945

/gene="bioc"

4190..4945

/gene="bioc"

/gene="bioc"

```

/codon_start=1
/transl_table=11
/protein_id="AAA23517.1"
/db_xref="GI:145427"
/translation="MATVVKQAIAAFGRAAAYEHQADLOROSADALLAML.PORKYT
HYLDGCGPGWMSRHREHAQVATLALDSPMLVOAROKDAADHYLAGDIESLPLATA
TFDLAWSNI.AVOWCGNLSTALRELYRVVRPGVVAFTLVGSLPERHOAAQVDEBP
HANRFLPDETEIOSLNGVYOHIOPTLTFMDALSAARSLKIGATHLHGGRPRLL
TRSOQRQLAMPQOQGRYPLTYHLFLGVARE"
4938..5597
/gene="biod"
/locus_tag="biod"
/feature="biod"
/codon_start=1
/transl_table=11
/product="gethioiotin synthetase"
/protein_id="AAA23518.1"
/db_xref="GI:145428"
/translation="MSKRYEYGTDEFGKTVASCALLQAAKRAGRYTAGYKPVASGS
EKTEGLRNSDALAIQNRNSLQDYATVNPYFAEPSPHISAOEGRPESLVMASG
LRALEQOADMVLYEGAGMFTPLSDFTFPADWYQOLPVLIVYGVKLGCTNNHMLTA
OVIQHAGLTLAGVANDVTPPGKRAHEKMTITMTIRRCWERSPGIQLKIMKOPES
T"
BASE COUNT      1363 a      1554 c      1631 g      1245 t
ORIGIN      4626 bp upstream of HpaI site; 18 min on K-12 map.

Query Match      100.0%; Score 25; DB 2; Length 5793;
Best Local Similarity 100.0%; Pred. No. 0.00032;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1      gcaagtcacgcacgacatcgat 25
          |||
Db      2095      GCAGTGCATCCGACGATTCGAT 2119

RESULT      6
LOCUS      A38246      5872 bp      DNA
DEFINITION      Sequence 1 from Patent WO9408023.
ACCESSION      A38246
VERSION      A38246.1      GI:2294844
KEYWORDS
SOURCE
ORGANISM      Escherichia coli.
              Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
              Escherichia.
REFERENCE      1 (bases 1 to 5872)
AUTHORS      Birch, O., Brass, J., Fuhrmann, M. and Shaw, N.
TITLE      BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
JOURNAL      Patent: WO 9408023-A 1 14-APR-1994;
              LONZA AG (CH)
COMMENT
Other publication PL 308301 950724
Other publication CA 2145400 940414
Other publication AU 4820293 940426
Other publication HU 71781 960228
Other publication SK 42085 951108
Other publication CZ 9500809 950913
Other publication FI 951547 950331
Other publication JP 8501694T 960227.
location/Qualifiers
1..5872
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/clone="PBO30A-15/9"
1..96
/function="PROMOTER PTAC"
/evidence=experimental
23..28
45..49
/standard_name="PROMOTER PTAC"
/evidence=experimental

RBS      105..119
/standard_name="BIOC RBS NO.9"
/evidence=experimental
117..1157
/gene="BIOB"
117..1157
/gene="BIOB"
/transl_table=11
/codon_start=1
/transl_table=11
/number=1
/evidence=experimental
/product="BIOTIN SYNTHASE"
/protein_id="CAA02325.1"
/db_xref="GI:2294845"
/translation="NAHRPRWTLQVTELEFERPLDLLEAQQVHROHFRPQVST
LSTIKTGACPEDCKYCPQSSRYKTEGLEARLEVEQVLESARAKAKASTRECMGAAM
KNPHERMDMPYLEQWGVKAMGLEAMGTEGTSQESQARLANAGIDYNNHNDPSPEF
YCNLTITRTYORBLDTLEKVRDAGIKVCSCGIVGGEYVKRAGILLDLANIPTTPES
VPINMVYKKGPIADNDVDVDFDTRTIANRIMPTSYKLSGRQNMBOQVAMC
PAGASITFGCKLITTPPEEDKDLFLKILGLNPQOTAVIAGDNEQOORLEQALMT
PQTDENYNAAL"
2284..2297
/standard_name="BIOC RBS"
2295..3050
/gene="BIOC"
2295..3050
/gene="BIOC"
/function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
/codon_start=1
/transl_table=11
/number=3
/product="PROTEIN"
/protein_id="CAA02326.1"
/db_xref="GI:2294846"
/translation="MATVVKQAIAAFGRAAAYEHQADLOROSADALLAML.PORKYT
HYLDGCGPGWMSRHREHAQVATLALDSPMLVOAROKDAADHYLAGDIESLPLATA
TFDLAWSNI.AVOWCGNLSTALRELYRVVRPGVVAFTLVGSLPERHOAAQVDEBP
HANRFLPDETEIOSLNGVYOHIOPTLTFMDALSAARSLKIGATHLHGGRPRLL
TRSOQRQLAMPQOQGRYPLTYHLFLGVARE"
3742..3752
/standard_name="BIOA RBS"
3750..5039
/gene="BIOA"
3750..5039
/gene="BIOA"
/standard_name="S-ADENOSYL-L-METHIONINE:
8-AMINO-7-OXONONANOATE AMINOTRANSF."
/EC_number="2.6.1.62"
/codon_start=1
/transl_table=11
/number=5
/evidence=experimental
/product="DAPA SYNTHASE"
/protein_id="CAA02327.1"
/db_xref="GI:2294847"
/translation="MTTDLAFDORHIMHPYTSPTLPYVYVSAAGCELIISDGR
LVDCSSMMAAHHGXNHPOLNAARKSOIDAMSHVMRGITHARAIELCRKYAMTQOP
LECFELADSGSAVEVAMKALQYWKAKGKARORPLTRNGYTGDTFGASVCDPDS
MHSIMKYTLPENLFAPARQSRMDGEMDERWVGARLMAHREHIAVILIEPTVOAG
TGGYMTLSATLTTRREVAETLSNCGACGFMGPTFMGKLPACAAANSIALIESGDWQ
QVADIEVQLREOLAPARDAEMVADVAVYLGAIGVETTHPYNMAALQKFEFGQVWIRP
FGKLIYMPPIYILPQOLRLTAAVNNAVDEFFPQ"
5088..5100
/standard_name="ORF1 RBS"
5098..5574
/gene="ORF1"
5098..5574
/gene="ORF1"
/function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
/codon_start=1
/transl_table=11
/number=6

```

```
/evidence=experimental
/product="PROTEIN"
/protein_id="CAA02328.1"
/db_xref="GI:2294848"
/translation="MKLSNDLRDDDKLPHRHVFGMGVDDGNISPHLAMDDVPAGTK
SFVVTCYDADPAGSGMHWVYNI.PADTRVLPQFGSLVAMPQGVQTRDPCKTG
YDGAAPPKGETHRIFYTVALDIERIDVDEGASGAMVGFNHFSLSASTVAMS"
5583..5605
5583..5644
stem_loop
terminator
/standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
TERMINATOR"
BASE COUNT      1318 a 1552 c 1695 g 1307 t
ORIGIN
Query Match      100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.00032;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1 gccagtgatcgccagcatcgcattcgat 25
      200 GCAGGTGCATCGCCAGCATTTGCAT 224
RESULT 7
A38251      5872 bp      DNA      PAT      05-MAR-1997
LOCUS      Sequence 6 from Patent WO9408023.
DEFINITION A38251
ACCESSION A38251
VERSION A38251.1 GI:2294849
KEYWORDS
SOURCE      Escherichia coli.
ORGANISM      Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (bases 1 to 5872)
AUTHORS      Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
TITLE      BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
JOURNAL      Patent: WO 9408023-A 6 14-Apr-1994;
      LONZA AG (CH)
COMMENT      other publication PL 308301 950724
      other publication CA 2145400 940414
      other publication AU 4820293 940426
      other publication HU 71781 960228
      other publication SK 42095 951108
      other publication CZ 9500809 950913
      other publication FI 951547 950331
      other publication JP 8501694T 960227.
      Location/Qualifiers
FEATURES
Source
1..5872
location/Qualifiers
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/clone="PBO30A15-9"
1141..1156
/standard_name="BIOF RBS"
1154..2308
/standard_name="BIOF"
1154..2308
/standard_name="BIOF"
1154..2308
/standard_name="BIOF"
/standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
/EC_number="2.3.1.47"
/codon_start=1
/transl_table=11
/number=2
/evidence=experimental
/product="KAPA SYNTHASE"
/protein_id="CAA02329.1"
/db_xref="GI:2294850"
/translation="MSWQKINAADARADALRRYPVAGAGRWLVADROYLNF
SSNDYLGSHFPIIRAWQQAEGFGISGGSGHYSGT SYVHQAEELEALMIGCSRA
LFIISGFANAVIAAMAKEDRIADRIASHLSLEASLSPSQLRRRAHNDVTHLAR
LIASPCGQOMVVEEGFVSMGDSAPLAEIDOVTOQHNGWMLVDDAHGTGVIGEGRG
```

```
SCWLOKYPPELLVTFGKFGVSGAAVLCSTVADYLLQFARHLIYSTSNPQAQAL
RASLAVRSDGDRREKLAALITRPAGVQDDLPFTLADSCSAIOPLIVGDNRSRALQ
AEKLQOQGWTAIRPTVPAGTARLRLLTITAHENQDDIRLLEVLHNGN"
3030..3045
/standard_name="BIOB RBS"
3043..3753
/standard_name="BIOB"
3043..3753
/standard_name="BIOB"
3043..3753
/standard_name="BIOB"
/standard_name="DETHIOBIOTIN SYNTHASE"
/EC_number="6.3.3.3"
/codon_start=1
/transl_table=11
/number=4
/evidence=experimental
/product="PTB SYNTHASE"
/protein_id="CAA02330.1"
/db_xref="GI:2294851"
/translation="MSKRYFVTGDTVEYKTVASCALLQAAKAGRTAGYKPPVAGS
KTEPEGLRNSDALALQRRNSSLQDYATVNPYFAEPTSPHLSAQGRPIESLVMSAG
LRALRQADWVLEVGAGGWTPLSDTFFADWVQEOQLPVILVGVKLCINHAMLTA
OYIOHAGLTIGWVANDVTPPGKRRAEYMTTLTRMIPAPLIGIPLMIAENPENATGK
YINLAFVDASTLIGFTSRL"
BASE COUNT      1318 a 1552 c 1695 g 1307 t
ORIGIN
Query Match      100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.00032;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1 gccagtgatcgccagcatcgcattcgat 25
      200 GCAGGTGCATCGCCAGCATTTGCAT 224
RESULT 8
A93674      5872 bp      DNA      PAT      22-JAN-2000
LOCUS      Sequence 1 from Patent EP0798384.
DEFINITION A93674
ACCESSION A93674
VERSION A93674.1 GI:6741862
KEYWORDS
SOURCE      Escherichia coli.
ORGANISM      Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (bases 1 to 5872)
AUTHORS      Birch,O. and Brass,J.
TITLE      Biotechnological method of producing biotin
JOURNAL      Patent: EP 0798384-A 1 01-OCT-1997;
      LONZA AG (CH)
FEATURES
Source
1..5872
location/Qualifiers
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/clone="PBO30A-15/9"
1..96
/promoter
/function="PROMOTER PTAC"
/evidence=experimental
-35_signal
-10_signal
45..49
/standard_name="PROMOTER PTAC"
/evidence=experimental
105..119
/standard_name="BIOB RBS NO.9"
/evidence=experimental
117..1157
/standard_name="BIOB"
117..1157
/standard_name="BIOB"
117..1157
/standard_name="BIOB"
/codon_start=1
```

```

/transl_table=11
/evidence=experimental
/product="Biotin SYNTHASE"
/protein_id="CAB69590.1"
/db_xref="GI:6741863"
/translation="MAHRPRTLQVTELEFPEKPLDLLFPAQVYHROHFPPOVOST
LSTKTCGCEPCDCYCCOSSRYKGLAEELMEVEYULESARAKAGSTRFCMGAW
KNRHRMPLEOVOCVAKAMGLEACHTGLTSQAKRLANAGLDYNNHNDTSEF
YGNITTTTIOERLDLEKVDAGIKVCSGIVGLCEVTRAGLILLOLANLPTPES
VPIMLVKVGTPLADNDVDADFEITIVAVRIMPTSVYRLSAGEQNEQOTAMC
FMAGANSIFYGCKLTTPNEPEDKDLDFKRLGILNPQGTAVLAGDNEQOURLQALMT
PDTDEYYNAAL"
2284..2297
/standard_name="BIOC RBS"
2295..3050
/gene="BIOC"
2295..3050
/gene="BIOC"
/function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
/codon_start=1
/transl_table=11
/product="PROTEIN"
/protein_id="CAB69591.1"
/db_xref="GI:6741864"
/translation="MATVKKQATIAAAGRAAAHYPOHADLQROSADALLAMIPKRYT
HYLDACGCGMSSRHRERHAYTALDLSPMIVQAKROKADADHYLAGDIESPLATA
TFDLASNLAWOMCGNLSTALRELYVRPKGVATTTIVGCSLPELHOAWQVDRP
HANRFLPDEIRSLINGVYCHHIOPIITLTFDALSAMSLGIGATLHESRDPRL
TRSOLORIOLAMQOQGRPLIYHLGLVIAHE"
3742..3752
/standard_name="BIOA RBS"
3750..5039
/gene="BIOA"
3750..5039
/gene="BIOA"
/standard_name="S-ADENOSYL-L-METHIONINE:8-AMINO-7-
OXONANONATE AMINOTRANSF."
/ec_number="2.6.1.62"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="DAPA SYNTHASE"
/protein_id="CAB69592.1"
/db_xref="GI:6741865"
/translation="MTTDLADQRHIMHPYTSMTSPVYVVSASGCELLISDGR
LVDGSSWMAALHGVNHPOLNAKMSQIDAMSHVMFGGITAPATELCRLIVAMPPOP
LECVLADSGSYAVEAMKALOYQWAGEARQRLTPFRNGVHGDTFGAMVCDDNS
MHSILKGLPELLEFAPQSRMDGEMDERDVGAFRLMAHRHEIAVITTEP YVQAG
GMRVHPEMIKRIKICDRREGILLADIATGFGTGLACGEAEIADPILCGKAL
TGGWMTLSNTLTTRVAVTISNGAGCTHGGFTENGPNLACAAANASIALLESQWQ
QVADIEVLEQELAPARDAENVADRVLGATGVETTHPVMAALOKKFFVQGWIRP
FEKLIYLPPIITLPEQDLRLIAANRAVODETFEQ"
5088..5100
/standard_name="ORFI RBS"
5098..5574
/gene="ORFI"
5098..5574
/gene="ORFI"
/function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="PROTEIN"
/protein_id="CAB69593.1"
/db_xref="GI:6741866"
/translation="MKLISNDLRDGRKLPHRHVPNGMGYDGDNTSPHLAMDVPAGTK
SFVVTGYDPAPGTGSMWHMWVYVLPADTRVLPGFGSGLVAMPDGVLYQTRTPGKRG
YDGAAPRKGETHRIETVHALDIRIVDGSAGSAMGVFVHHISLASASITAMFS"
5583..5605
/standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
TERMINATOR"
1552 c 1695 g 1307 t

```

```

ORIGIN
Query Match 100.0%; Score 25; DB 9; Length 5872;
Best Local Similarly 100.0%; Pred. No. 0.00032;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 gcagtgcatgcgcacattcgat 25
Db 200 GCAGTGCAATGCCACATTTCGAT 224

RESULT 9
A93679 5872 bp DNA PAT 22-JAN-2000
LOCUS Sequence 6 from Patent EP0798384.
DEFINITION A93679
ACCESSION A93679
VERSION A93679.1 GI:6741867
KEYWORDS
SOURCE
ORGANISM
Escherichia coli.
Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
1 (bases 1 to 5872)
Birch, O. and Brass, J.
Biotecnological method of producing biotin
patent: EP 0798384-A 6 01-OCT-1997;
LONZA AG (CH)
FEATURES
location/Qualifiers
1..5872
/organism="Escherichia coli"
/strain="DSM498"
/db_xref="taxon:562"
/clone="PB030A15-9"
1141..1156
/standard_name="BIOF RBS"
1154..2308
/gene="BIOF"
1154..2308
/gene="BIOF"
/standard_name="8-AMINO-7-OXONANONATE SYNTHASE"
/ec_number="2.3.1.47"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="KAPA SYNTHASE"
/protein_id="CAB69594.1"
/db_xref="GI:6741868"
/translation="MSMOERKINALDARAADALRRRYPAAGAGRTVADROYLNF
LFIISGPAANOVIAMAKEDRIADRSHASLSLEASLSFQLRFRFANHVTHLAR
LLASPCGQOVVTEGVFSMDGSDAPLARIQOVYQDQNSMLVDAHGVGIEGQGRG
SCWLOKVRPELVITFGKGVGGAVALICSTVADYILOFAHRLYSTSMPPAQAOL
RASLAVIRSDGEGDAREKLAALITFRRAQVODLPFLADSCAIOPLIYDMSRALQL
AEKLRQCGCWVTAIRPPTVAGTARLRLLTAAHEQDIDRLLEVLHGNG"
3030..3045
/standard_name="BIOD RBS"
3043..3753
/gene="BIOD"
3043..3753
/gene="BIOD"
/standard_name="DETHIOBIOTIN SYNTHASE"
/ec_number="6.3.3.3"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="DTR SYNTHASE"
/protein_id="CAB69595.1"
/db_xref="GI:6741869"
/translation="MSKRYFYNGTPEVKGTVASCALLQAKAAGYTRAGTKPPVAGS
KTRPGRNSDALAQRNSQIOLDYATVNPPTFAPTSPHITSADEGRPIESLVSAG
LRAIRQADWVTVBGAGMFTPLSDTFFADWVQOQLPVILVGVKLGCTNHMLTA
QVIOHAGLTLAGVANDVTPPGKRHAEEYMTLTTRMIPAPRLGLGELFWLAENBNATGK

```

```

BASE COUNT 1318 a 1552 c 1695 g 1307 t

```


BASE COUNT 1318 a 1552 c 1695 g 1307 t
 ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 5872;
 Best Local Similarity 100.0%; Pred. No. 0.00032;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 gcaggtgcatgccagcattcgat 25
 Db 200 GCAGGTGCATGCCAGCATTTTCAT 224

RESULT 10
 LOCUS ARI01809 5872 bp DNA PAT 14-FEB-2001
 DEFINITION Sequence 1 from patent US 6083712.
 ACCESSION ARI01809
 VERSION ARI01809.1 GI:12812607
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 5872)
 AUTHORS Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
 TITLE Biotechnological method of producing biotin
 JOURNAL Patent: US 6083712-A 1 04-JUL-2000;
 FEATURES Location/Qualifiers
 source 1..5872

BASE COUNT 1318 a 1552 c 1695 g 1307 t
 ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 5872;
 Best Local Similarity 100.0%; Pred. No. 0.00032;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 gcaggtgcatgccagcattcgat 25
 Db 200 GCAGGTGCATGCCAGCATTTTCAT 224

RESULT 11
 LOCUS ARI01810 5872 bp DNA PAT 14-FEB-2001
 DEFINITION Sequence 6 from patent US 6083712.
 ACCESSION ARI01810
 VERSION ARI01810.1 GI:12812608
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 5872)
 AUTHORS Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
 TITLE Biotechnological method of producing biotin
 JOURNAL Patent: US 6083712-A 6 04-JUL-2000;
 FEATURES Location/Qualifiers
 source 1..5872

BASE COUNT 1318 a 1552 c 1695 g 1307 t
 ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 5872;
 Best Local Similarity 100.0%; Pred. No. 0.00032;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 gcaggtgcatgccagcattcgat 25
 Db 200 GCAGGTGCATGCCAGCATTTTCAT 224

RESULT 12
 LOCUS AE000180
 DEFINITION Escherichia coli K12 MG1655 section 70 of 400 of the complete genome.

ACCESSION AE000180 U00096
 VERSION AE000180.1 GI:1786988
 KEYWORDS
 SOURCE Escherichia coli K12.
 ORGANISM Escherichia coli K12.
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.

REFERENCE 1 (bases 1 to 11022)
 AUTHORS Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V., Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F., Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J., Mau,B., and Shao,Y.
 TITLE The complete genome sequence of Escherichia coli K-12
 JOURNAL Science 277 (5331), 1453-1474 (1997)
 MEDLINE 9742617
 PUBMED 9278503

REFERENCE 2 (bases 1 to 11022)
 AUTHORS Blattner,F.R.
 TITLE Direct Submission
 JOURNAL University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
 Email: ecolig@genetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459

REFERENCE 3 (bases 1 to 11022)
 AUTHORS Blattner,F.R.
 TITLE Direct Submission
 JOURNAL University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
 Email: ecolig@genetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459

REFERENCE 4 (bases 1 to 11022)
 AUTHORS Plunkett,G. III.
 TITLE Direct Submission
 JOURNAL submitted (13-OCT-1998) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
 COMMENT This sequence was determined by the E. coli Genome Project at the University of Wisconsin-Madison (Frederick R. Blattner, director). Supported by NIH grants HG00301 and HG01428 (from the Human Genome Project and NCHGR). The entire sequence was independently determined from E. coli K12 strain MG1655. Predicted open reading frames were determined using Genemark software, kindly supplied by Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA.

30332 [e-mail: mark@amr.gatech.edu]. Open reading frames that have been correlated with genetic loci are being annotated with CG site nos., unique ID nos. for the genes in the E. coli Genetic Stock Center (CGSC) database at Yale University, kindly supplied by Mary Berlyn. A public version of the database is accessible (http://cgsc.biology.yale.edu). Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful by correlating it with other data. Comments to the authors are appreciated. Updated information will be available at the E. coli Genome Project's World Wide Web site (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and its annotations are periodically updated: this is version M54. No sequence changes. Annotation updates: updated gene identifications and products; all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'b' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names.

FEATURES
 source Location/Qualifiers
 1..11022
 /organism="Escherichia coli K12"
 /strain="K12"
 /sub_strain="MG1655"

repeat_region
/db_xref="taxon:83333"
/note="REP (repetitive extragenic palindromic) element;
contains 4 REP sequences"
complement(147..1430)
/gene="ybhc"
/note="b0772"
complement(147..1430)
/gene="ybhc"
/function="orf: Not classified"
/note="f427: 98 pct identical to fragment YBHC_ECOLI
SM:P46130 (300 aa) but contains 127 additional C-terminal
residues"
/codon_start=1
/transl_table=1
/product="putative pectinesterase"
/protein_id="AAC73859.1"
/db_xref="GI:1786989"
/translation="MTFVSRLALAFGVTLTACSTPPDQPSDQAPGSSRI
LSAEQNPDAOHFASLTPGAAMNPSPITIPADPFVGVAGTCTHTT1QAAVD
AAIKRNRKROYIAVMGEYGVYPAACGTLTGTEKPLDVKIGSLDGMSPA
LGDVYAGNRPAYALRTDQVQIINNVLIGKNTPEFVNSGVNRLTNPRTLV
NSTIEGDVIVSGAVEDNTEFRVNVNRTQEAFAATLSNITYGFLAVNSFN
AFEDGVAQLGRSLDVANTNGVYIRDSAINEGFNPAKPAWDAVISNRPAGTGSVD
DNDEIQNLNDINYNRMEXNNRGSKVVAEAK"
complement(1507..1535)
/note="factor Sigma70: predicted +1 start at 806574"
complement(1582..2058)
/gene="ybhb"
/note="b0773"
complement(1582..2058)
/function="orf: Unknown"
/note="f158: 99 pct identical to YBHB_ECOLI SM: P12994"
/codon_start=1
/transl_table=1
/product="orf, hypothetical protein"
/protein_id="AAC73860.1"
/db_xref="GI:1786990"
/translation="MKLISNDLRDCKLPHRFVNGMGVGDNISPLAMDVPACTK
SEVVTCPDPAITGSGMHWYVNLPAIDRVLPQSGSIVAMPDGVLOTRTDEKGTG
YGAAPKGETHRVYFTVAHLIERIDDEGSGAMGVNHFHSLASASTTAMS"
complement(2108..2124)
/note="central position to predicted promoter:85.5"
/bound_moiety="Rnas predicted site"
complement(2117..3406)
/gene="bioA"
/note="b0774"
complement(2117..3406)
/EC_number="2.6.1.62"
/function="enzyme: Biosynthesis of cofactors, carriers:
Biotin"
/note="f429: 100 pct identical to BIOA_ECOLI SM: P12995"
/codon_start=1
/transl_table=1
/product="7.8-diaminopelargonic acid synthetase"
/protein_id="AAC73861.1"
/db_xref="GI:1786991"
/translation="MTTDDLAFDORNIWHPYTSPLPYVVSAGCELLISDGR
LVDSMSWMAIHGYNHPOLNAAMKSOIDASHVFGGITHAPAIETCRIVAMTPD
LCEVFLADSGVAEAVAMKALQYWAQGEAROFLEPFGKCHDPTGANSVCDDPS
MHSIMKGYLLENLAPAPOSRMDGEMBERGVARLMAAHRHETIAVITPTVQAG
GMRYHPBEMLKRIKICDREGILIDETATGERTGLFACETAEIADPILDGLKAL
TGCTMTSATITTEVAETISNAGAGCGFPMGPNPLACAANASIALLESQMOO
QVADIEVLKEQLAPADAEVADVRTGAIGVETTHPVMMAALQKFEQGVWIRP
FGKLIIMPYIILPOLQRLTAAVNVAODETFEQ"
complement(2193..2221)
/gene="bioA"
/note="factor Sigma70: predicted +1 start at 807260"
3411..3450
/note="central position to bioB promoter: -20"

protein_bind
/bound_moiety="bioB documented site"
3411..3450
/note="central position to predicted promoter: -20;genetic
evidence for the site"
/bound_moiety="bioB predicted site"
complement(3411..3450)
/note="central position to predicted promoter:50"
/bound_moiety="bioB predicted site"
/note="central position to bioA promoter:50"
/bound_moiety="bioB documented site"
3413..3441
/note="factor Sigma70: promoter bioB; documented +1 at
808525"
complement(3447..3473)
/note="factor Sigma70: promoter bioA; documented +1 at
808515"
3493..4533
/gene="bioB"
/note="b0775"
3493..4533
/gene="bioB"
/EC_number="2.8.1.-"
/function="enzyme: Biosynthesis of cofactors, carriers:
Biotin"
/note="o346: 99 pct identical to BIOB_ECOLI SM: P12996"
/codon_start=1
/transl_table=1
/product="biotin synthesis, sulfur insertion?"
/protein_id="AAC73862.1"
/db_xref="GI:1786992"
/translation="MAHRPMTLSQVTELEFEKPLDLLFPAQGVHROHPDPROVYST
LLSRTGACPEDECKYCPQSSRYKTGLAEERLMEYVLESARAKAGSFRPGMAW
KNPHERMPYLEQWQGVKAMGLACACTLIGLSQOYLRANGLDYNNIDTSPER
YGNITTRTVOERLDTEKVDAGIKVCSGYICGTVDYRAGLLQLNLTPTPES
VPIMLVKVGCTPLADNDADPFIITIVARIIMPTSTVRLSAGREONNEGTOAMC
FMAGANSIFVGCKLITTPNEBEDKDLQFLKGLINPQTVAVLGDNEQDRLBOALMT
PDDEYTNAAAL"
4530..5684
/gene="bioF"
/note="b0776"
4530..5684
/gene="bioF"
/EC_number="2.3.1.47"
/function="enzyme: Biosynthesis of cofactors, carriers:
Biotin"
/note="o384: 100 pct identical to BIOF_ECOLI SM: P12998"
/codon_start=1
/transl_table=1
/product="8-amino-7-oxononanoate synthase"
/protein_id="AAC73863.1"
/db_xref="GI:1786993"
/translation="MSWOKRINAALDARRAADALRRRYPVAGAGRWLVADDROYLNF
SSNDYGLSHHPQIRAWOQGAEOFGIGSGSGHVSGYVVAQALFEELAEVLGSR
LFIISGFANQAVIAIAMAEDRIADRLSHASLEASLSPSOLRREAFHNVHTLAR
LHSPCGOQNVATEGVFMDGDSAPLAEIQOVTQOONWMLWDDAHGCVIGEGRG
SCWLQKVPRELLVTRGKGFVSGAANVLCSSIVADYLLQFAHLLYTSMPAQOAL

Query Match 100.0%; Score 25; DB 1; Length 11022;
Best Local Similarity 100.0%; Pred. No. 0.0003;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gcagtgacgcgcacgattcgat 25
|||||
Db 3576 GCAAGTGCATCGGCACATTTCAT 3600
|||||

RESULT 13
AE005258
LOCUS AE005258 13501 bp DNA
DEFINITION Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 82
of 155
ACCESSION AE005258 AE005174

```

VERSION      AE005258.1 GI:12513751
KEYWORDS
SOURCE       Escherichia coli O157:H7 EDL933.
ORGANISM     Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
              Escherichia.
REFERENCE    1 (bases 1 to 13501)
AUTHORS      Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
              Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
              Postel,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
              Grothbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
              Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
              Welch,R.A. and Blattner,F.R.
              Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
              Nature 409 (6819), 529-533 (2001)
TITLE        Nature 409 (6819), 529-533 (2001)
JOURNAL      21074935
MEDLINE      11206551
PUBMED
REFERENCE    2 (bases 1 to 13501)
AUTHORS      Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
              Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
              Postel,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
              Grothbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
              Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
              Welch,R.A. and Blattner,F.R.
              Direct Submission
              Submitted (22-OCT-2000) Laboratory of Genetics, University of
              Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
              Location/Qualifiers
              1. 13501
              /organism="Escherichia coli O157:H7 EDL933"
              /strain="EDL933"
              /serotype="O157:H7"
              /db_xref="taxon:155864"
              /note="enterohaemorrhagic"
              <1. 7576
              /note="O-island #36. Region of the EDL933 chromosome not
              homologous to E. coli K-12 MG1655; cryptic prophage
              CP-933K; includes one copy of the 13 bp direct repeat that
              flanks the prophage"
              66. 665
              /gene="lomK"
              /note="20981"
              66. 665
              /gene="lomK"
              /function="putative membrane; Other or unknown (Phage or
              prophage Related)"
              /note="Residues 1 to 199 of 199 are 72.86 pct identical to
              residues 1 to 199 of 199 from Genpept 118 :
              g117532789|gb|AA63231.1|AF151091_2 (AF151091) lom
              [prophage P-ElbA]"
              /codon_start=1
              /transl_table=1
              /product="putative outer membrane protein of prophage
              CP-933K"
              /protein_id="AAG55137.1"
              /db_xref="GI:12513752"
              /translation="MRKYCAILSAICLAVSGVPAMASEHOSITAGYLHASTDAPG
              SDDLNGIWKRYEFTDILGLITSPSTANADECKTYSOTRMHEDYVRRRMSTWAG
              PSVRNWEFSAVAMAGVAYSRVSTFSGDYPTVNKRKTHDVLITGSDARTSNTSLAM
              GAGVQNPTEESVAVDVAYEXSGSDWRTDFIVGYK"
              719. 2041
              /gene="20982"
              719. 2041
              /gene="20982"
              /function="putative structure; Structural component (Phage
              or Prophage Related)"
              /note="Residues 164 to 440 of 440 are 68.79 pct identical
              to residues 381 to 645 of 645 from Genpept 118 :
              g114585436|gb|AAD25464.1|AF125520.59 (AF125520) putative
              tail fiber protein [bacteriophage 933w]"
              /codon_start=1
              /transl_table=1
              /product="putative tail component of prophage CP-933K"

gene
CDS
/protein_id="AAG55138.1"
/db_xref="GI:12513753"
/translation="MGMAAYOISGVLKDKGACKPIONCTIOLKARNSTXYVVNTVASE
NDEGRTISMDEYTCGYVTLLVREFPSPSHACTISYVEDSOPGLNDFLAMEDEAR
PALRFRFEQVEEARHAEAEKKNAGEETARNAKASISAKASAAANAPTSADSE
SAROAESAASAKSEASSSASAAKASESIQSATDALSKKTESAAGNARARA
TTTEKARESAASQSAEQSRTAEDAVNR1PTVVGPGPGGPGPGPGKDGKE
KDGTPGATGTGRTGPGPGPGGRTGRTGRTGRTGRTGRTGRTGRTGRTGRTG
GRTGRTGRTGRTGRTGRTGRTGRTGRTGRTGRTGRTGRTGRTGRTGRTGRTG
TGTFTLDPKDATQVGMGMHQLVRRGDDGPMQDVGKGLDVGSDTRTGE"
2004. 2312
/gene="20984"
2004. 2312
/gene="20984"
/function="orf; Other or unknown (Phage or Prophage
Related)"
/note="Residues 1 to 102 of 102 are 98.03 pct identical to
residues 28 to 129 of 129 from Genpept 118 :
g114585437|gb|AAD25465.1|AF125520.60 (AF125520)
hypothetical protein [bacteriophage 933w]"
/codon_start=1
/transl_table=1
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55139.1"
/db_xref="GI:12513754"
/translation="MRKAVIOAELENDNNILKRLKQSLCGCGKHDCEGNSLTAOLR
LGPADLLESDENCIIPEDORVLTQVYLLDADKKQIQCVVRPLQILRADGTWENIGMK"
2489. 3469
/gene="20985"
2489. 3469
/gene="20985"
/function="orf; Other or unknown (Phage or Prophage
Related)"
/note="Residues 13 to 326 of 326 are 55.73 pct identical
to residues 18 to 331 of 336 from Genpept 118 :
g116960367|gb|AA633527.1| (AF170176) hypothetical protein
predicted by Glimmer [Salmonella typhimurium LT2]"
/codon_start=1
/transl_table=1
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55140.1"
/db_xref="GI:12513755"
/translation="MISPIRTTFHNSVNIYVSSPCOTVSPACKETELAKYDEKPTPLF
KLEPFENPERYKDEVPYVNTKOHPIYLDNVTNNAIRISDRMIGTFVGDSTVNQKRAF
S KLEPFENVMYKIREVDVDFSMYDBKLSDIYHDIICORLREDEKRDREYLLNLEELR
EISKAQDSLSIMYAKKRNHAMPDEFERLALLKAGEIIFCYNTKKNHGISFGEGCIYLD
MDMLITGLGTTTPDSTISNHYVDRNDNSVNIENSATITVNSNHPALLEGISFMHKKVD
AHPIYDGLGKGVKAYFNFPTLHNINHCDFIERNHRIINTSQTCCSM"
3530. 4522
/gene="20986"
3530. 4522
/gene="20986"
/function="orf; Other or unknown (Phage or Prophage
Related)"
/note="No significant matches"
/codon_start=1
/transl_table=1
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAG55141.1"
/db_xref="GI:12513756"
/translation="MKIPSLDSNPNFSPAGCYSAPIAPNRAENAYADVLDIGKRIFL
SAADLSNVEYSIVRAVHDSRSRLIDQHTVDMIGTIVDALSRSGTPEFADNSYGIINER
VHIGCIKRYNEIELNEESSVKRIDIQSLTLCLEIYVGOEPIPTICEAENENPERX
VSFSAADPTDSEYMSQEGLIHETIHHVYTSOSPSSDSEIETLGPTEILARKVAOELG
WSVPDFKGYAEPREARHLRLNLNALRQAARHENEERAFERLGTISDYVEASPDFT
EYSAVSNIGYGFITQHDPPGLAINDNLQDANQIOLYHGAPYITFFGVDVKNHOD"
5350. 6231
/gene="20989"
5350. 6231
/gene="20989"
/function="orf; Other or unknown (Phage or Prophage
Related)"

```

```

/note="No significant matches"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAC55142.1"
/db_xref="GI:12513757"
/transl_table=11
/translation="MISPSYVNLGCSWNSITRLTSPDRNVLSVDAAVHSDNGAUV
KVGNRTRYVATONKPCVTRSHSGCTTLLHRLGMPKGEISKRLEMLNASPVSAKM
FRGTVNSRDLPLVDAPPELPSVDNRLISVGNVIGKGNAAVVEDADATKVLK
FTTSQSEVYSEVRCFNOYVAGASAKITGNGDIIIGIRMDINGESLISLPAQ
AEHATYDFRLEKQKILLFVDTTETVLVDRANKERNPIDISSYVSDRSMSQIMQ
SYHGKODLSIVLSKI"
/length=7160
/gene="20990"
/size=6462..7160
/gene="20990"
/function="orf; Other or unknown (Phage or Prophage
Related)"
/note="No significant matches"
/codon_start=1
/transl_table=11
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AAC55143.1"
/db_xref="GI:12513758"
/transl_table=11
/translation="MRPSTLNLVHLHSSRSSMSDTEISLVKASSVQMIKNNQRLP
QGDHNIYDQIEALDKIGSTETGRVLLNATISIRLSEYVILHNSRLCYAHRD
IDENHRTGSDPHCNLNAVEYPCGGISVDFHATIVFHEHLHFNHNLNGERLKEVS
SRAESQKSYLLLEAFRTVGLAFSEVISEKHEIKERPRSTYPSDALIHDDNT
VSLGFOOVRHLPII"
/complement(7668..8144)
/gene="ydhb"
/note="20992"
/complement(7668..8144)
/gene="ydhb"
/function="orf; Unknown function"
/note="Residues 1 to 158 of 158 are 99.36 pct identical to
MG1655: B0773"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AG55144.1"
/db_xref="GI:12513759"
/transl_table=11
/translation="MKLISNDLRDQKLPHRHNVNGMGYDGNISPHLANDVPAGTK
SEVVCYDPDAPATGSGMMWVYVNLPAOTRVLPQSGSLVAMPDGLVOTRPGKTC
IDGAPPKGEGHRYLFTVHALDVERIDVEDGASGAMGVNHFHSLASATIMFS"
/complement(8203..9492)
/gene="DioA"
/note="20993"
/complement(8203..9492)
CDS
Query Match 100.0%; Score 25; DB 1; Length 13501;
Best Local Similarity 100.0%; Pred. No. 0.0003;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gcaggtgcacgcacgacattcgat 25
|||||
Db 9662 GCAGGTGCATCCGACGACATTTCGAT 9686

RESULT 14
AP002553 297816 bp DNA BCT 07-MAR-2001
LOCUS Escherichia coli O157:H7 DNA, complete genome, section 4/20.
ACCESSION AP002553 BA000007
VERSION AP002553.1 GI:13360211
KEYWORDS
SOURCE
Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RIMD 0509952)
DNA.
Escherichia coli O157:H7
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE
1 (sites)

AUTHORS
Makino, K., Yokoyama, K., Kubota, Y., Yutsudo, C. H., Kimura, S.,
Kurokawa, K., Ishii, K., Hattori, M., Tatsuno, I., Abe, H., Iida, T.,
Yamamoto, K., Ohnishi, M., Hayashi, T., Yasunaga, T., Honda, T.,
Sasakawa, C. and Shinagawa, H.
Complete nucleotide sequence of the prophage WY7-Sakai carrying the
verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
derived from the Sakai outbreak
Genes Genet. Syst. 74 (5), 227-239 (1999)
2 (sites)
Ohnishi, M., Murata, T., Nakayama, K., Kuhara, S., Hattori, M.,
Kurokawa, K., Yasunaga, T., Yokoyama, K., Makino, K., Shinagawa, H. and
Hayashi, T.
Comparative analysis of the whole set of rRNA operons between an
enterohemorrhagic Escherichia coli O157:H7 strain MG1655
Escherichia coli K-12 strain MG1655
Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
3 (sites)
Yokoyama, K., Makino, K., Kubota, Y., Matanabe, M., Kimura, S.,
Yutsudo, C. H., Kurokawa, K., Ishii, K., Hattori, M., Abe, H., Iida, T.,
Yamamoto, K., Hayashi, T., Yasunaga, T., Honda, T., Sasakawa, C. and
Shinagawa, H.
Complete nucleotide sequence of the prophage WY1-Sakai carrying the
Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
O157:H7 strain derived from the Sakai outbreak
Gene 258 (1-2), 127-139 (2000)
4 (sites)
Hayashi, T., Makino, K., Ohnishi, M., Kurokawa, K., Ishii, K.,
Yokoyama, K., Han, C.-G., Ohtsubo, E., Nakayama, K., Murata, T.,
Tanaka, M., Tobe, T., Iida, T., Takami, H., Honda, T., Sasakawa, C.,
Ogasawara, N., Yasunaga, T., Kuhara, S., Shiba, T., Hattori, M. and
Shinagawa, H.
Complete genome sequence of enterohemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12
DNA Res. 8 (1), 11-22 (2001)
5 (bases 1 to 297816)
Ohnishi, M., Kurokawa, K., Makino, K., Yasunaga, T., Shinagawa, H. and
Hayashi, T.
Direct Submission
Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
Information Research Center, 3-1, Yamadaoka, Suita, Osaka 565-0871,
Japan (E-mail: ken@gen-info.osaka-u.ac.jp,
URL: http://www.gen-info.osaka-u.ac.jp/, Tel: 81-6-6879-8365,
Fax: 81-6-6879-2047)
genome project.
COMMENT
FEATURES
Location/Qualifiers
1..297816
/organism="Escherichia coli O157:H7"
/strain="O157:H7"
/sub_strain="RIMD 0509952"
/db_xref="taxon:83334"
79..1245
/gene="Ecs0753"
79..1245
/gene="Ecs0753"
/note="similar to succ_ECOLI g11786948 percent identity
100 in 388 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence="not-experimental"
/product="succinyl-CoA synthetase beta subunit"
/protein_id="BAB34176.1"
/db_xref="GI:13360212"
/transl_table=11
/translation="NNLHVEYQAKOLFAPRYGLPAPYVACTTPREAFASRTGAGPW
VKCOVHAGGRGKAGYKVVNSKEDIAPAFEMKRIYVYQYDANQGVNOLVBAAT
DIKELIYAVVDRSRVPMASRSGEYIEKVAEETPHLHKVADLPGLGPMYDQ
RELAFRLGEGKIVQDFTKIFMGLATPFLERDLALLETINPLVITKQGLDLDKRLGA
DGNALRQDPLREMRDQSEDPREAQAWMLNVAALDNGICGVNAGSLAMGMDIY
KIHGEGPANFLVDVGATKERYTEAFKIIISDPKAVAVLNI FGIVYCDLIAGGIIG
AAVEGVNPPVVVRLEGNAEIGARRKLADSGINTIAKGLIPDAQGVAAVEGK"

```

gene
1245. .2114
/gene="ECS0754"
CDS
1245. .2114
/note="similar to SUCD_ECOLI g11786949 percent identity
100 in 289 aa (Conserved in E.coli K-12)"
/codon_start=1
/transl_table=11
/evidence=not-experimental
/product="succinyl-CoA synthetase alpha subunit"
/protein_id="BAB34177.1"
/db_xref="GI:13360213"
/translation="MSILDKNTKVTYCGFTGSGCTFHSEQAIAVYGRKMGVTPGK
GTHLGLPVNTFVREAVATGASVIVPAFCDSILEAIDAGIKLIITTEGIP
LMTVTKVKLDAGVGMIGPNCGVITTEGICGPHIRKPGVIGVSGTLYE
AKQOTDYEGCGSTCGTGGTIDILEMEKDPOTEALVIMGEIGSAEEA
AAIKHVTKPVVGTIAGVTAKKGRMGHAGAIINAGGCTADENFALAEAGVTVS
LADIGALKTVLK"
2781. .3686
/gene="ECS0755"
CDS
2781. .3686
/gene="ECS0755"
/note="probable transcriptional regulator, similar to
transcriptional regulators e.g. glycine cleavage system
transcription activator (gcv operon activator)"
Escherichia coli g11417043|sp|P32064|GCV_A_ECOLI percent
identity 31 in 300 aa"
/codon_start=1
/transl_table=11
/evidence=not-experimental
/product="putative transcription regulator"
/protein_id="BAB34178.1"
/db_xref="GI:13360214"
/translation="MRGKIPEKTELVTFEVVAHESYTRAEELALTSAAVFRVYSAL
EEFIRPLFSHKKRIEINDTKGYIVKTEINKLERDNTIMTQPVQVTELVN
PTSTHMLINLHEPTKIHDDIIVNTHSLANNEDREYDAVIMENECAPMAVEY
LFEETILPVCSSILAMSDOKLSVALLTEPLIHSTRTGMEEFALSGVSPVYN
NGPREFDLSMLIAVRSNLGVALPREPAILOHDDSDMYIPCDVPIRTONRRTMWE
EKSDSPHLOQFREMILAKSVVQEM"
complement(3720. .4322)
/gene="ECS0756"
CDS
complement(3720. .4322)
/gene="ECS0756"
/note="probable cob(I)alamin adenosyltransferase, similar
to cob(I)alamin adenosyltransferases (corrinoid
adenosyltransferases) e.g. [Escherichia coli]
g11151481|sp|P130401|BTUR_ECOLI percent identity 67 in 200
aa"
/codon_start=1
/transl_table=11
/evidence=not-experimental
/product="putative cob(I)alamin adenosyltransferase"
/protein_id="BAB34179.1"
/db_xref="GI:13360215"
/translation="MEARISTEHHVORQOKLEQVDTRVAALIEKKGIILVFTGNGK
KSTAFGVTRAVGHGKTGVGAOYIKGOMNGEYNILOPGEVHHIMGTFTETONR
QADIDAKEVWSESKRMADKRYDLVLDLTYMLAHYHIDTEVLAISLGNRAQOSV
IVTGRGCHSQTILKMAVDVSEIRPVKHAHFNKIQAPQIDW"
complement(4332. .5984)
/gene="ECS0757"
CDS
complement(4332. .5984)
/gene="ECS0757"
/note="probable fumarate hydratase, similar to fumarate
hydratases e.g. fumarate hydratase class I, aerobic
(fumarate) - Escherichia coli
g11120598|sp|P00923|FUMA_ECOLI percent identity 68 in 545
aa"
/codon_start=1
/transl_table=11
/evidence=not-experimental
/product="putative fumarate hydratase"
/protein_id="BAB34180.1"
/db_xref="GI:13360216"

translation="MSKRFITWQPEPLNKKDQETELISDQHTITVELDGEVYIAKE
ALTLLARAPFASFLSALQOVALSTINDQOASNDKAYVALLOLLMAEVSARAVLP
NCODPDTATVASKSGQOIMTGNQDEALSKITYTPOENNLRSQNALPMYEVNTO
TMIPAOIDISAVAGDEYHFLVKNKGGSANALYQETKSLDPEKATLALIKMSYL
GTAACPPIHIAFVVGSLASDOTLKIAKLAISKRYVDNLPTSGNEGQAFROIIEKRYL
EASOOGFIAOGFGCKYFAHDRIKRLPHEGSCPIAMALSCSADRINKAIINRGLWL
EKLEHNPQGYIPASIRENHQHOVLNDRLRPMODLALPYGTGVSSJGPVIVAR
DIAHAKIKARLDSGEPMEPEYKHHIVVYAGCAPENNACSLGPTTGSGMDGYIDF
QAAGSLVMLSKGNSQOVTDCRKHGCFNIGSGAAILAOEYVSLSCLEPELPG
MEAVMMMEVEENIPAEILVDDKGNFFSOFEOQHRKASCPAGH"
complement(6092. .7372)
/gene="ECS0758"
CDS
complement(6092. .7372)
/gene="ECS0758"
/note="probable transport protein, similar to
glutamate/aspartate transport proteins (proton glutamate
symport proteins) e.g. [Bacillus stearothermophilus]
g1121467|sp|P24943|GLT_BACST percent identity 38 in 416
aa, also similar to C4-dicarboxylate transport"
/codon_start=1
/transl_table=11
/evidence=not-experimental
/product="putative transport protein"
/protein_id="BAB34181.1"
/db_xref="GI:13360217"
/translation="MKKISLTMMILALVIGMIIGVINMTASPEAKLVAQESIFT
TIFRLIKMTIAPLVSTLVGIAKKGDAAKALFEFTSLFICASLSTALGLITV
NFMPTGGINFVAGHAEPTGVAAAREPTIKVFTISNAPFIVDAHANEIIOIVFSI
FLCCSLPAIGESKSAIVHALDSLAHMLKLTGVVWLPATVFAAISALIERGLAV
VSGILFMGEYFTFLLWLILGLALIVYGPICIRLTRLSPALAFTSSSEAP
GLEKLEQEGVSPRIASFVLPIDGYSNVLGSMAYCSFAVITQACNTHISGRITM
LLIIMLSKMGAVPRASVVIATLNOINPEAGIILMGVDPDLDMGRSNTVMSN
AMGAAVSRWEGEFGCEGCKALKPNESVALP"
complement(7533. .7850)
/gene="ECS0759"
CDS
complement(7533. .7850)
/gene="ECS0759"
/note="unknown"
/codon_start=1
/transl_table=11
/evidence=not-experimental
/product="hypothetical protein"
/protein_id="BAB34182.1"

Query Match 100.0%; Score 25; DB 2; Length 297816;
Best Local Similarity 100.0%; Pred. No. 0.00022;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gcaggtgcacgcacgacatccgat 25
|||||
Db 92394 GCAGTGCATCCCGACGATTCGAT 92418

RESULT 15
PSEHRPIA/C 3603 bp DNA BCT 21-OCT-1993
LOCUS Pseudomonas syringae hrpJ genes, complete cds.
DEFINITION L11582
ACCESSION L11582
VERSION L11582.1 GI:151267
KEYWORDS hrpJ gene; protein translocation.
SOURCE Pseudomonas syringae (strain 61) DNA.
ORGANISM Pseudomonas syringae
Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
Pseudomonas
1 (bases 1 to 3603)
REFERENCE
AUTHORS Huang, H.-C., Xiao, Y., Lin, R.-H., Lu, Y., Hutcheson, S.W. and
Collmer, A.
TITLE Characterization of the Pseudomonas syringae pv. syringae 61 hrpJ
and hrpJ genes: homology of hrpJ to a superfamily of proteins
associated with protein translocation
Mol. Plant Microb Interact. 6, 515-520 (1993)
JOURNAL
FEATURES
source Location/Qualifiers
1. .3603

AUTHORS Entcheva, P., Liebl, W., Johann, A., Hartsch, T. and Streit, W.R.
TITLE Direct cloning from enrichment cultures, a reliable strategy for isolation of complete operons and genes from microbial consortia
JOURNAL Appl. Environ. Microbiol. 67 (1), 89-99 (2001)
MEDLINE 11133432
PUBMED 20575196
REFERENCE 2 (bases 1 to 5526)
AUTHORS Entcheva, P., Liebl, W. and Streit, W.R.
TITLE Direct Submission
JOURNAL Submitted (31-MAR-2000) Mikrobiologie und Genetik, Universitaet Goettingen, Grisebachstr. 8, Goettingen 37077, Germany
FEATURES Location/Qualifiers
 1..5526
 /organism="uncultured bacterium pcosHE2"
 /db_xref="taxon:143797"
 /clone="pcosHE2"
 /note="unknown organism, cosmid clone derived from environmental consortium"
 complement(52..528)
 /note="ORF1"
 /codon_start=1
 /transl_table=11
 /product="hypothetical 17.1 kDa protein in modc-bioA intergenic region"
 /protein_id="AAG60577.1"
 /db_xref="GI:12620125"
 /translation="MKLISNDLDCDKLPHRHVFGMGYDGNISPHLADNDVPAATK SFVTCGDDAPDGTGSGMHVYVNLPAIDRYLPQCGSGLVAMPDGVLOTPTDCKAG YGCAAPKGETHRYIFTHALDIRIDVDEGASGAMGVNHFSLASASITAMS"
 complement(587..1876)
 /gene="bioA"
 /complement(587..1876)
 /gene="bioA"
 /note="7, 8-diaminopelargonic acid synthetase-aminotransferase"
 /codon_start=1
 /transl_table=11
 /product="DAPA-aminotransferase BioA"
 /protein_id="AAG60578.1"
 /db_xref="GI:12620126"
 /translation="MTTDLAFDORHIMHPYTSMTSPLPVYFVSAEGCELLSDGRK LNDGSMWAAIHGYNHNPOLNMAKMSQIDAMSHVFGGITHAPATELCCKLVAAMPPOK LKCYPLADSGSAVEYAKKALQYNOAKKEARPLTPNGYHGDPFGMSYCDSDG MSILWGYLPENLFAPAPOSKMDGDEWEDMGFAPALMAHHETIAVITIEIYOGAG GMRVYHEWLRKRIKICDEGILLADIADLGGRTGLFACHEAIIADIIICLCKAL TGGMTLSATLTREVAETISNGEACFMHGPFGNPLACAAANASIALLESQWQO QVADIEVOLREOLAPARDAEMVADVILCAIGVETTHVNNAAALOKFEVEGVNIRP FGLTIYMPYIILLPQOLQRLTAANRAVADDTFFCQ"
 1963..3003
 /gene="bioB"
 /1963..3003
 /gene="bioB"
 /codon_start=1
 /transl_table=11
 /product="biotin synthase BioB"
 /protein_id="AAG60579.1"
 /db_xref="GI:12620127"
 /translation="MAHRPRWTLSQVTELEFKEPLDLLLEAQQVHRQHPDPROVOST LLSIKTACPEDCKYCPDSSRYKTGLEAEFLMEVEVLESARAKAAGSTRICMGAW KNPERDMPTLEQWGVGKIDGLEACWTLGTLSAQORLANGLDYVHNLDTSPEF YGNLITRTYQERIDLEKVDAGIKVCSGIVGLGIVDRAGLILLOLANPTSPES VPIIMLVKVKSTPLADNDVDADPFTTIVARTIMPTSYVRLSAGREONBDQOAMC FMAGANSIFGCKLITTPNEBDKDLQFLKGLINPQOTAVLAGNDQOORLEQALMT PDDEYVNAAL"
 3000..4154
 /gene="bioF"
 /3000..4154
 /gene="bioF"
 /note="8-amino-7-oxononanoate synthase"
 /codon_start=1
 /transl_table=11
 /product="KAPA synthetase BioF"
 /protein_id="AAG60580.1"

gene /db_xref="GI:12620128"
 /translation="MSMQEKINNALDARRADALRRRYPVAGAGRMIVADROXYLNF SSNDYLGISHHPQIIRAMQOGAEQFGISGSGSHVSTYVHOLAEELAEMLGYSRA LLFISGANAVAVIAAMAKDRILAPORLSHASSILEASIS/PSOLRFAHNDYTHAR LLASPCGOALVVTGEGVSGMDSDAPLAEIDQVYQOQNGMLMDVDAHGTCVGEQSG SCWLOKVAPELLVYTGEGVSGAAVICSSTVADYLLQFARHLIYSTSGPPAQDAL RASLAVTRDEGDARREKLVLIAHFRAGVODDPTLADSCSAIOPLYVDNSRALQL AEKLQCGCWYTAIRPPVPACTARLRLTLAAHEMODIDRLLEVLHGNC"
 4141..4896
 /gene="bioC"
 /4141..4896
 /gene="bioC"
 /note="reaction step prior to synthesis of pimeloyl-CoA"
 /codon_start=1
 /transl_table=11
 /product="biotin biosynthesis protein BioC"
 /protein_id="AAG60581.1"
 /db_xref="GI:12620129"
 /translation="MATVKKOALAAAFGRAAALYEOHADLQROSAVDVLLAMPKRYT HYLDAGCGGGMTRHRRHAGVATLADSPMLVOAROKDAADHYLAGDIESPLATA TFDLANSNLAYQWGNLSLALBELYRVRSGVYAFPTLYOGSLPELHOAOVADEP HANRLPDEIEOSLNGVYHGHOTPTLNFMDALSAMRLKIGATHHIEGRPRL TRSOLQRLQAMPQOQGRPLYHLFGLVARE"
 4883..5526
 /gene="bioD"
 /4883..5526
 /gene="bioD"
 /note="DTB synthetase"
 /codon_start=1
 /transl_table=11
 /product="delhibiotin synthetase BioD"
 /protein_id="AAG60582.1"
 /db_xref="GI:12620130"
 /translation="MLVSKRYPTGDPREVECKTVASCALLQAKAAGRTAGYKPVAS GSKRTPEGLNSDALALQNRSSLDQDAYVNTYTAEPSPHISAOGRPLESSVMS SGLRALEQADWVLYVAGAGWETPLSDITTFADWYTOEQLPYLYVYKLCICINHAML TAOALIHAGLTLGAVNADVTPPGKRHAHYITTLRMIPAPILGELPWLAEPE"
BASE COUNT 1274 a 1507 c 1567 g 1178 t
ORIGIN

Query Match 72.0%; Score 18; DB 2; Length 5526;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 catgccagcattcgat 25
 |||||
Db 2053 CATGCCAGCATTCGAT 2070

RESUME 18
 G01446
 LOCUS Dm0466 Drosophila P1 library Drosophila melanogaster STS genomic clone DS07967 SP6, sequence tagged site.
 ACCESSION G01446
 VERSION G01446.1 GI:684849
 KEYWORDS STS
 SOURCE fruit fly.
 ORGANISM Drosophila melanogaster.
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 479)
REFERENCE Rubin, G.
AUTHORS Berkeley Drosophila Genome Project
TITLE Unpublished (1994)
JOURNAL
COMMENT Contact: Berkeley Drosophila Genome Project
 Primer A: CTCCTACTGGGACTTGCTA
 Primer B: CACCAAGGCATATCCGA
 STS size: 192

PCR Profile:
 Annealing: 58 degrees C PCR Cycles: 30
 Protocol:
 Template: p1 library Pools
 Primer: 1 uM each
 dNTPs: 250 uM each
 Tag Poly: 0.05 units/u1
 Total Vol: 15 u1

Buffer:
 MgCl2: 1.5mM
 KCl: 50 mM
 Tris-HCl: 50 mM
 pH: 8.3
 Gelatin: .001 %

The p1 library has been distributed to 16 regional sites. A list of these sites is available from Flybase, via anonymous ftp to flybase.indiana.edu in the file flybase/allied-data/genome-projects/1b1/LBL.doc.

FEATURES

source

1. .479
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone="DS07967"
 /note="Vector: PAD10Sact1; The p1 library was made by D. Smoller in D. Hartl's lab (see Smoller et al., Chromosome 100: 487). Chromosomal position was mapped by polytene chromosome in situ hybridization in the Hartl lab (see Hartl et al., PNAS 91: 6824). STS's were generated by sequencing the ends of the Drosophila insert in these p1 clones, by the W. Kimmerly, C. Martin, and M. Palazzolo lab at LBL."

STS
 primer_bind 113 a 120 c 97 g 145 t 4 others
 BASE COUNT complement(233. .250)
 ORIGIN

Query Match 68.0%; Score 17; DB 54; Length 479;
 Best Local Similarity 100.0%; Pred. No. 16;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 gcatgccagcatctcg 23
 ||||||||||||||||
 Db 396 GCATGCCAGCATTTTCG 412

RESULT 19
 AC019940/c
 LOCUS
 DEFINITION
 AC019940 53067 bp DNA HTG 03-JAN-2000
 Drosophila melanogaster, *** SEQUENCING IN PROGRESS *** in ordered
 pieces
 AC019940
 VERSION
 AC019940.1 GI:6664957
 KEYWORDS
 HTG; HTGS_PHASE2.
 SOURCE
 fruit fly.
 ORGANISM
 Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 I (bases 1 to 53067)
 Adams, M. and Venter, J.C.
 Direct Submission
 Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive,
 Rockville, MD, USA

REFERENCE

AUTHORS

TITLE

JOURNAL

Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive,
 Rockville, MD, USA
 This sequence was identified as CDL10211463 by the submitter.
 For more information on this record e-mail to fly@celera.com.
 * NOTE: This is a 'working draft' sequence.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.

FEATURES
 source
 Location/Qualifiers
 1. .53067
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"

BASE COUNT 14944 a 11879 c 11720 g 14524 t
 ORIGIN

Query Match 68.0%; Score 17; DB 65; Length 53067;
 Best Local Similarity 100.0%; Pred. No. 10;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 gcatgccagcatctcg 23
 ||||||||||||||||
 Db 13629 GCATGCCAGCATTTTCG 13613

RESULT 20
 AC013431
 LOCUS
 DEFINITION
 AC013431 153870 bp DNA HTG 31-JAN-2000
 Drosophila melanogaster chromosome X clone BACR22h11 (D1191)
 RPCI-98 22.H.11 map 13E-13F strain Y; cn bw sp, *** SEQUENCING IN
 PROGRESS *** 73 unordered pieces.

AC013431
 VERSION
 AC013431.8 GI:6838815
 HTG; HTGS_PHASE1.
 SOURCE
 fruit fly.
 ORGANISM
 Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 I (bases 1 to 153870)

REFERENCE
 AUTHORS
 Celniker, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazek, R.G.,
 Burenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,
 Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,
 Hinkle, A., Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K.,
 Kearney, L., Lee, B., Lewis, S., Li, P., Ling, H., Moshrefi, A.R.,
 Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S., Pfeiffer, B.,
 Richards, S., Sethi, H., Svirskas, R.R., Wan, K.H., Webster, D.,
 Woolley, P., Yang, S., Yee, M., Yu, C. and Rubin, G.M.
 Sequencing of Drosophila melanogaster
 Unpublished
 2 (bases 1 to 153870)

Celniker, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazek, R.G.,
 Burenhoff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,
 Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,
 Hinkle, A., Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K.,
 Kearney, L., Lee, B., Lewis, S., Li, P., Ling, H., Moshrefi, A.R.,
 Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S., Pfeiffer, B.,
 Richards, S., Sethi, H., Svirskas, R.R., Wan, K.H., Webster, D.,
 Woolley, P., Yang, S., Yee, M., Yu, C. and Rubin, G.M.
 Direct Submission
 Submitted (11-NOV-1999) Drosophila Genome Center, Lawrence Berkeley
 Laboratory, MS 64-121, Berkeley, CA 94720, USA

On Jan 31, 2000 this sequence version replaced gi:6532028.
 For further information about this sequence, including its location
 and relationship to other sequences, please visit our sequence
 archive Web site (<http://www.fruitfly.org/sequence/>) or send email
 to bdg@fruitfly.berkeley.edu. All contigs in this submission meet
 the following cutoffs: length >= 200 bases.

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 73 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 995: contig of 995 bp in length
 * 996 1075: gap of unknown length
 * 1076 1560: contig of 485 bp in length
 * 1561 1640: gap of unknown length
 * 1641 2230: contig of 590 bp in length


```

* 2231 2310: gap of unknown length
* 2311 2956: contig of 646 bp in length
* 2957 3036: gap of unknown length
* 3037 3621: contig of 585 bp in length
* 3622 3701: gap of unknown length
* 3702 4535: contig of 834 bp in length
* 4536 4615: gap of unknown length
* 4616 5559: contig of 944 bp in length
* 5560 6539: gap of unknown length
* 6540 6615: contig of 896 bp in length
* 6616 7660: gap of unknown length
* 7661 7740: gap of unknown length
* 7741 8730: contig of 990 bp in length
* 8731 8810: gap of unknown length
* 8811 9404: contig of 594 bp in length
* 9405 9484: gap of unknown length
* 9485 10725: contig of 1241 bp in length
* 10726 10805: gap of unknown length
* 10806 11578: contig of 773 bp in length
* 11579 11658: gap of unknown length
* 11659 12834: contig of 1176 bp in length
* 12835 12914: gap of unknown length
* 12915 13501: contig of 587 bp in length
* 13502 13581: gap of unknown length
* 13582 14462: contig of 881 bp in length
* 14463 15566: contig of 1024 bp in length
* 15567 15646: gap of unknown length
* 15647 16366: contig of 720 bp in length
* 16367 16446: gap of unknown length
* 16447 17301: contig of 855 bp in length
* 17302 17381: gap of unknown length
* 17382 18390: contig of 1009 bp in length
* 18391 18470: gap of unknown length
* 18471 19409: contig of 939 bp in length
* 19410 19489: gap of unknown length
* 19490 20726: contig of 1237 bp in length
* 20727 20806: gap of unknown length
* 20807 22052: contig of 1246 bp in length
* 22053 22132: gap of unknown length
* 22133 23015: contig of 883 bp in length
* 23016 23095: gap of unknown length
* 23096 24622: contig of 1527 bp in length
* 24623 24702: gap of unknown length
* 24703 26689: contig of 1987 bp in length
* 26690 26770: gap of unknown length
* 26771 27825: contig of 1056 bp in length
* 27826 27905: gap of unknown length
* 27906 29904: contig of 1999 bp in length
* 29905 29984: gap of unknown length
* 29985 31588: contig of 1604 bp in length
* 31589 31668: gap of unknown length
* 31669 33112: contig of 1444 bp in length
* 33113 33192: gap of unknown length
* 33193 33744: contig of 2182 bp in length
* 33745 35454: gap of unknown length
* 35455 37223: contig of 1769 bp in length
* 37224 37303: gap of unknown length
* 37304 38978: contig of 1675 bp in length
* 38979 39058: gap of unknown length
* 39059 41337: contig of 2279 bp in length
* 41338 41417: gap of unknown length
* 41418 42461: contig of 1044 bp in length
* 42462 42541: gap of unknown length
* 42542 45823: contig of 3282 bp in length
* 45824 45903: gap of unknown length
* 45904 47247: contig of 1344 bp in length
* 47248 47327: gap of unknown length
* 47328 50153: contig of 2826 bp in length
* 50154 50233: gap of unknown length
* 50234 52182: contig of 1949 bp in length
* 52183 52262: gap of unknown length

```

FEATURES

source

```

* 52263 56756: contig of 4494 bp in length
* 56757 56836: gap of unknown length
* 56837 63525: contig of 6689 bp in length
* 63526 63605: gap of unknown length
* 63606 71274: contig of 7669 bp in length
* 71275 71354: gap of unknown length
* 71355 78141: contig of 6787 bp in length
* 78142 78221: gap of unknown length
* 78222 86570: contig of 8349 bp in length
* 86571 86570: gap of unknown length
* 86571 96227: contig of 9577 bp in length
* 96228 96307: gap of unknown length
* 96308 107432: contig of 11124 bp in length
* 107433 107511: gap of unknown length
* 107512 121618: contig of 14107 bp in length
* 121619 121698: gap of unknown length
* 121699 139797: contig of 18099 bp in length
* 139798 139877: gap of unknown length
* 139878 140388: contig of 511 bp in length
* 140389 140469: gap of unknown length
* 140470 141011: contig of 543 bp in length
* 141012 141092: gap of unknown length
* 141093 141593: contig of 422 bp in length
* 141594 142057: contig of 464 bp in length
* 142058 142137: gap of unknown length
* 142138 142690: contig of 553 bp in length
* 142691 142770: gap of unknown length
* 142771 143344: contig of 574 bp in length
* 143345 143424: gap of unknown length
* 143425 143857: contig of 433 bp in length
* 143858 143937: gap of unknown length
* 143938 144603: contig of 666 bp in length
* 144604 144683: gap of unknown length
* 144684 145116: contig of 433 bp in length
* 145117 145196: gap of unknown length
* 145197 145734: contig of 538 bp in length
* 145735 145814: gap of unknown length
* 145815 146224: contig of 410 bp in length
* 146225 146304: gap of unknown length
* 146305 146859: contig of 555 bp in length
* 146860 146939: gap of unknown length
* 146940 147585: contig of 646 bp in length
* 147586 147665: gap of unknown length
* 147666 148081: contig of 416 bp in length
* 148082 148161: gap of unknown length
* 148162 148539: contig of 378 bp in length
* 148540 148619: gap of unknown length
* 148620 148950: contig of 331 bp in length
* 148951 149030: gap of unknown length
* 149031 149375: contig of 345 bp in length
* 149376 149455: gap of unknown length
* 149456 149833: contig of 398 bp in length
* 149834 149933: gap of unknown length
* 149934 150532: contig of 599 bp in length
* 150533 150612: gap of unknown length
* 150613 151147: contig of 535 bp in length
* 151148 151227: gap of unknown length
* 151228 151577: contig of 350 bp in length
* 151578 151657: gap of unknown length
* 151658 151964: contig of 307 bp in length
* 151965 152044: gap of unknown length
* 152045 152629: contig of 585 bp in length
* 152630 152709: gap of unknown length
* 152710 153245: contig of 536 bp in length
* 153246 153325: gap of unknown length
* 153326 153870: contig of 545 bp in length.

```

Location/Qualifiers

1..153870

/organism="Drosophila melanogaster"

/strain="y; cn bw sp"

/db_xref="taxon:7227"

/chromosome="X"

Query Match 68.0%; Score 17; DB 63; Length 153870;
 Best Local Similarity 100.0%; Pred. No. 9.5;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 7 gcatgccgcatcttcg 23
 |||||||||||||||
 Db 63673 GCATGCCGCACTTCG 63689

RESULT 21
 AC037435 183937 bp DNA HTG 15-JAN-2001
 LOCUS Mus musculus chromosome 11 clone RP23-9813 map 11, WORKING DRAFT
 DEFINITION
 AC037435
 AC037435.2 GI:12229535
 HTG: HTGS_PHASE1; HTGS_DRAFT.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 183937)
 Birren, B., Linton, J., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
 Birren, B., Linton, J., Nusbaum, C., Lander, E.,
 Mus musculus chromosome 11, clone RP23-9813
 2 (bases 1 to 183937)
 Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,
 Boguslavsky, L., Boukhalter, B., Brown, A., Burkett, G.,
 Campion, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
 Collymore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S.,
 Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D.,
 Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
 Grand-Pierre, N., Grant, G., Hagos, B., Heatford, A., Horton, L.,
 Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
 Klein, J., Larocque, K., Lamazares, R., Landers, T., Lehoczy, J.,
 Levine, R., Liu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
 McCarthy, M., McEwan, P., McGuirk, A., McKernan, K., McPheeters, R.,
 Melidim, J., Menus, L., Miho, T., Miranda, C., Mienga, V., Morrow, J.,
 Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
 O'Neill, D., Oliver, T.M., Oliver, C., Peterson, K., Pierre, N.,
 Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
 Stange, R., Thoman, N., Stojanovic, N., Subramanian, A., Talamas, J.,
 Testa, S., Theodore, J., Tirrell, A., Travers, M., Trigg, J.,
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
 Young, G., Zainoun, J., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (08-APR-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www.seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 Project Information
 Center project name: 18882
 Center clone name: 98_T3

Summary Statistics
 Sequencing vector: M13; M77815; 4% of reads
 Sequencing vector: Plasmid; n/a; 96% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 177644 bases at least Q40
 Consensus quality: 177644 bases at least Q30
 Consensus quality: 180041 bases at least Q20
 Insert size: 188000; agarose-fp
 Insert size: 181637; sum-of-configs

Quality coverage: 4.7 in Q20 bases; agarose-fp
 Quality coverage: 4.9 in Q20 base.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 24 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 23881: contig of 23881 bp in length
 23882 23981: gap of 100 bp
 23982 24469: contig of 488 bp in length
 24470 24569: gap of 100 bp
 24570 24755: contig of 186 bp in length
 24756 24855: gap of 100 bp
 24856 25997: contig of 1142 bp in length
 25998 26097: gap of 100 bp
 26098 26764: contig of 667 bp in length
 26765 26864: gap of 100 bp
 26865 28426: contig of 1562 bp in length
 28427 28526: gap of 100 bp
 28527 30488: contig of 1962 bp in length
 30489 30588: gap of 100 bp
 30589 32514: contig of 1926 bp in length
 32515 32614: gap of 100 bp
 32615 36175: contig of 3561 bp in length
 36176 36275: gap of 100 bp
 36276 38734: contig of 2459 bp in length
 38735 38834: gap of 100 bp
 38835 40927: contig of 2093 bp in length
 40928 41027: gap of 100 bp
 41028 43453: contig of 2426 bp in length
 43454 43553: gap of 100 bp
 43554 46744: contig of 3191 bp in length
 46745 46844: gap of 100 bp
 46845 50369: contig of 3525 bp in length
 50370 50469: gap of 100 bp
 50470 56799: contig of 6330 bp in length
 56800 56899: gap of 100 bp
 56900 57617: contig of 718 bp in length
 57618 57717: gap of 100 bp
 57718 68517: contig of 10800 bp in length
 68518 68617: gap of 100 bp
 68618 82793: contig of 14176 bp in length
 82794 82893: gap of 100 bp
 82894 96371: contig of 13478 bp in length
 96372 96471: gap of 100 bp
 96472 112208: contig of 15737 bp in length
 112209 112308: gap of 100 bp
 112309 127004: contig of 14696 bp in length
 127005 127104: gap of 100 bp
 127105 147474: contig of 20370 bp in length
 147475 147574: gap of 100 bp
 147575 174126: contig of 26552 bp in length
 174127 174226: gap of 100 bp
 174227 183937: contig of 9711 bp in length.

FEATURES
 source
 1..183937
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /chromosome="11"
 /map="11"
 /clone="RP23-9813"
 /clone_1lb="RP23-9813 Female Mouse BAC"
 1..23881
 /note="assembly-fragment
 clone_end:SP6
 vector_side:left"
 23982..24469
 /note="assembly-fragment"
 24570..24755
 /note="assembly-fragment"

```

misc_feature      24856..25997
                    /note="assembly_fragment"
misc_feature      26098..26764
                    /note="assembly_fragment"
misc_feature      26865..28426
                    /note="assembly_fragment"
misc_feature      28527..30488
                    /note="assembly_fragment"
misc_feature      30589..32514
                    /note="assembly_fragment"
misc_feature      32613..36175
                    /note="assembly_fragment"
misc_feature      36276..38734
                    /note="assembly_fragment"
misc_feature      38835..40927
                    /note="assembly_fragment"
misc_feature      41028..43453
                    /note="assembly_fragment"
misc_feature      43554..46744
                    /note="assembly_fragment"
misc_feature      46845..50369
                    /note="assembly_fragment"
misc_feature      50470..56799
                    /note="assembly_fragment"
misc_feature      56900..57617
                    /note="assembly_fragment"
misc_feature      57718..68517
                    /note="assembly_fragment"
misc_feature      68618..82793
                    /note="assembly_fragment"
misc_feature      82894..96371
                    /note="assembly_fragment"
misc_feature      96472..112208
                    /note="assembly_fragment"
misc_feature      112309..127004
                    /note="assembly_fragment"
misc_feature      127105..147474
                    /note="assembly_fragment"
misc_feature      147575..174126
                    /note="assembly_fragment"
misc_feature      174227..183937
                    /note="assembly_fragment"
misc_feature      clone_end::T7
                    vector_side:right"
BASE COUNT      47977 a 43626 c 42743 g 47258 t 2333 others
ORIGIN

```

```

Query Match      68.0%: Score 17: DB 71: Length 183937;
Best Local Similarity 100.0%: Pred. No. 9.4: Indels 0: Gaps 0:
Matches 17: Conservative 0: Mismatches 0:

```

```

OY      6      tgcacgcacgacatc 22
          |||||
Db      90965 TGCATGCCACGACATTTC 90981

```

```

RESULT 22
AE003500      327446 bp      DNA      INV      04-OCT-2000
LOCUS      Drosophila melanogaster genomic scaffold 142000013386053 section 17
DEFINITION      of 30 complete sequence.
ACCESSION      AE003500 AE002593
VERSION      AE003500.2 GI:10728273
KEYWORDS      HTG.
SOURCE      fruit fly.
ORGANISM      Drosophila melanogaster
              Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
              Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
              Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE      1 (bases 1 to 327446)
              Adams,M.D., Celinker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
              Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galie,R.F.,

```

```

George, R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
Brandon,R.C., Rogers,I.H., Blazer,R.G., Champ,M., Pfeiffer,B.D.,
Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
Miklos,G.L., Abri,J.F., Agbayan,A., An,H.J.,
Andrews-Pfankoch,C., Baldwin,D., Balley,R.M., Basu,A.,
Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
Botchan,M.R., Bouck,J., Brokstein,P., Brothier,P., Burtis,K.C.,
Busam,D.A., Butler,H., Cadieu,C., Davenport,L.B., Davies,P., de
Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
Podson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
Dunn,P., Durbin,K.J., Evangelista,C.C., Ferrara,C., Ferreira,S.,
Fleischmann,W., Foster,C., Gabrielian,A.E., Gary,N.S.,
Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorelli,J.H., Gu,Z.,
Guan,P., Harris,M., Harris,N.L., Harvey,D., Helman,T.J.,
Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
Kennison,J.A., Kerchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
Krautiz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
Morris,D., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
Nusskern,D.R., Paclet,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kimios,I.,
Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
Stapleton,M., Strong,R., Sun,E., Sytkas,R., Tector,C., Turner,R.,
Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wasserman,D.A.,
Weinstock,G.M., Weissbach,J., Williams,S.M., Woodage,T.,
Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,C.J.
The genome sequence of Drosophila melanogaster
Science 287 (5461), 2185-2195 (2000)
20196006
TITLE      2 (bases 1 to 327446)
JOURNAL    Adams,M.D., Celinker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
REFERENCE  Direct Submission
AUTHORS    Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
JOURNAL    Rockville, MD, USA
COMMENT    On Oct 9, 2000 this sequence version replaced gi:7293108.
FEATURES   location/Qualifiers
           source
           1..327446
           /organism="Drosophila melanogaster"
           /db_xref="taxon:7227"
           /chromosome="X"
           complement(join(772..903,968..1504,1572..2077,2153..2821,
           4145..>4380))
           /gene="CG9281"
           /product="CT26402"
           /db_xref="FLYBASE:FBan0009281"
           /db_xref="FLYBASE:FBgn0030672"
           complement((772..>4380))
           /gene="CG9281"
           /map="13E12-13E14"
           /db_xref="FLYBASE:FBan0009281"
           /db_xref="FLYBASE:FBgn0030672"
           complement(join(772..903,968..1504,1572..2077,2153..2813))
           /gene="CG9281"
           /note="CG9281 gene product"
           /codon_start=1
           /db_xref="FLYBASE:FBan0009281"
           /db_xref="FLYBASE:FBgn0030672"
           /protein_id="AAFA48493.1"
           /db_xref="GI:7293109"
           /translation="MSNAKKREAKKKDANKTKIOIPSTKANGVPERLFEKTL
           CAKLEEARISAEARSCTGLAVHPRSDVITANFSTFPESSSLDQTMVEYDERITL
           GILGNGCKSLILAVLIGREVPPHIDTIRFIPASSKSLDQVMEYDERITL
           EKLAELAMSEEDDAQEOLIDIERLDMDSADLAEVKAARITLIGIDPKAWQOQKQAD

```

[illegible]

Query Match	68.0%;	Score 17;	DB 4;	Length 327446;
Best Local Similarity	100.0%;	Pred. NO. 8.9;		
Matches 17;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
0y	7	gcataccagcattcg 23		
Db	70808	GCATCGCCAGCATTCG 70824		
RESULT 23				
AF146614/c				
LOCUS				
DEFINITION	AF146614	1186 bp	DNA	BCT
				09-AUG-1999
				Erwinia carotovora subsp. carotovora xerD (xerD) gene, partial cds,
				Dsbc precursor (dsbc) gene, complete cds; and RecJ (recJ) gene,
				partial cds.
ACCESSION	AF146614			
VERSION				
KEYWORDS	AF146614.1	GI:5712695		
SOURCE				
ORGANISM				
				Pectobacterium carotovorum subsp. carotovorum.
				Pectobacterium carotovorum subsp. carotovorum
				Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
				Pectobacterium.
REFERENCE	1 (bases 1 to 1186)			
AUTHORS	Vincent-Sealy,L.V., Thomas,J.D., Commander,P. and Salmond,G.P.			
TITLE	Erwinia carotovora Dsba mutants: evidence for a periplasmic-stress			
	signal transduction system affecting transcription of genes			
	encoding secreted proteins			
JOURNAL	Microbiology 145 (Pt 8), 1945-1958 (1999)			
MEDLINE	99392457			
REFERENCE	2 (bases 1 to 1186)			
AUTHORS	Thomas,J.D.			
TITLE	Direct Submission			

[illegible]

JOURNAL Submitted (29-APR-1999) Biological Sciences, University of Warwick,
Gibbet Hill Road, Coventry CV4 7AL, UK

FEATURES

source 1..1186
/organism="Pectobacterium carotovorum subsp. carotovorum"
/strain="SCRI193"
/sub_species="carotovora"
/db_xref="taxon:555"
<1..189
/gene="xerD"
/gene="xerD"
/gene="xerD"
/note="similar to Escherichia coli xerD
integrase/recombinase in SwissProt Accession Number
P21891"
/codon_start=1
/transl_table=11
/evidence="not_experimental"
/product="XerD"
/protein_id="AAD47611.1"
/db_xref="GI:5712697"
/translation="KLSPLVLRHAFATHLINHGAGFTGRTDALGAQRSEPHADLHPCR
DGAETRASAAHVRHRT"
222..1031
/gene="dsbc"
222..227
/gene="dsbc"
246..251
/gene="dsbc"
258..264
/gene="dsbc"
/note="putative"
315..377
/gene="dsbc"
/note="putative sec-dependent export signal"
315..1031
/gene="dsbc"
/function="periplasmic disulfide bond formation"
/note="disulfide isomerase; similar to Erwinia
chrysanthemi and Escherichia coli dsbc proteins in
SwissProt Accession Numbers P39691 and P21892"
/codon_start=1
/transl_table=11
/product="dsbc precursor"
/protein_id="AAD47610.1"
/db_xref="GI:5712696"
/translation="MKKGLDLLSLVATATNFADDAIAIKQTIAASGYARGGDPAFA
DGRMKTIVTDSGLYISDGAFAARPALRRERTMPNTNKLILKDLALQDMIV
KAAQEKIVTFTSITGCGYCHLHEQKDYNALGITVRYLAIFRQCKSPAAKDMQSI
MCVADRKAAPDSAMKGDALSAATCTKDIAHYDLGIQGVGTPALIVLDDGLVPGYQ
GPKELAMLEAKRSKKTGC"
378..1028
/gene="dsbc"
/product="dsbc"
669..680
/gene="dsbc"
/note="region: putative CPHC motif typical of the active
sites of disulfide isomerases"
1044..>1186
/gene="recJ"
1044..>1186
/gene="recJ"
/note="theoretical protein: similar to Erwinia
chrysanthemi and Escherichia coli RecJ single-stranded
DNA-specific exonucleases in SwissProt Accession Numbers
P39693 and P21893"
/codon_start=1
/transl_table=11
/evidence="not_experimental"
/product="RecJ"
/protein_id="AAD47612.1"
/db_xref="GI:5712698"
/translation="MITQLRRRLAEDIDLPTVPLRLRLYAHGVKGAQOELESLIS

BASE COUNT

295 a 315 c 323 g 253 t

ORIGIN

Query Match 64.0%; Score 16; DB 2; Length 1186;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 gcacgccagcatttc 22
|||||

Db 993 GCATGCCAGCATTTC 978

RESULT 24

AK024397/c

LOCUS
DEFINITION
Homo sapiens cDNA FLJ14335 fis, clone PLACE4000411, highly similar
to Homo sapiens mRNA; cDNA DKFp586D0624 (from clone
DKFZ586D0624).

ACCESSION

AK024397.1 GI:10436775

VERSION

011000 capping; fis (full insert sequence).

KEYWORDS

SOURCE

Homo sapiens placenta cDNA to mRNA, clone_11b:PLACE4
clone:PLACE4000411.

ORGANISM

Homo sapiens

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

```

Query Match      64.0%: Score 16; DB 89; Length 4674;
Best Local Similarity 100.0%: Pred. No. 49;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 ggtgcatcgccagcat 19
|||||
Db 2446 GGTGCATCGCCAGCAT 2431

RESULT 25
AF206717.c 11910 bp DNA SYN 12-SEP-2000
LOCUS AF206717 Shuttle vector p13 hypothetical proteins, HtrU (htru), Ardu (ardu),
DEFINITION RepU (repu), and Resu (resu) genes, complete cds.
ACCESSION AF206717
VERSION AF206717.1 GI:7262998
KEYWORDS
SOURCE Shuttle vector p13.
ORGANISM Shuttle vector p13
REFERENCE 1 (bases 1 to 11910)
AUTHORS Meima,R. and Lidstrom,M.E.
TITLE Characterization of the minimal replicon of a cryptic Deinococcus
radiodurans SARK plasmid and development of versatile Escherichia
coli-D. radiodurans shuttle vectors
APPL. Environ. Microbiol. 66 (9), 3856-3867 (2000)
MEDLINE 20422197
PUBMED 10966401
REFERENCE 2 (bases 1 to 11910)
AUTHORS Meima,R. and Lidstrom,M.E.
TITLE Direct Submission
JOURNAL Submitted (18-NOV-1999) Chemical Engineering, University of
Washington, Benson Hall, PO Box 351750, Seattle, WA 98195, USA
FEATURES
source location/Qualifiers
1..11910
/organism="Shuttle vector p13"
/specific_host="Deinococcus radiodurans SARK"
/db_xref="taxon:115340"
/note="derived from pUE10"
105..593
/note="orf-a"
/codon_start=1
/transl_table=11
/evidence="not_experimental"
/product="hypothetical protein"
/protein_id="AAF44052.1"
/db_xref="GI:7263004"
/translation="MGQIVBGGDDVPPNDGQODVHRLPLFVGFVKDKFAMLRG
GKERONDPEIDTSWPKRMVVEAFHPDDGGGVRAGGAEGRGFSMPDVLAPKEGDEHG
EQLDLVLAELREVGEAGOLGQDIDGRVQPSRVSRTSPSPAGNCRVLRMGQKRPK
AV"
misc-feature complement(257..683)
/note="5 mismatches with sequence deposited in Genbank
Accession Number M94966"
649..653
599..1078
/note="OrfB"
/codon_start=1
/transl_table=11
/evidence="not_experimental"
/product="hypothetical protein"
/protein_id="AAF44047.1"
/db_xref="GI:7262999"
/translation="MRFVAFDGSSEPSQALKAQITLGAAGAELEVVTVEPPVSG
FVQWAGLNPDYREMLEQONARAOQTOTLIGEGVQGRVOTLGRPIEVLTEAHAV
DLVVGTHGYRGMDRIILGSVTEFLTRAEVPLLVYR"
1187..2436
/gene="htru"
1187..1192
/gene="htru"
1198..2436
/gene="htru"
1198..2436
/gene="htru"
CDS

```

```

/codon_start=1
/transl_table=11
/product="Htru"
/protein_id="AAF44048.1"
/db_xref="GI:7263000"
/translation="MKTAMPILTLILLACGTDPESTRTQTTSTRTATNFPAGGA
ARDSSSPGACGTASPDGRLDYDONTIDVYKTDQGVFVTRPDGSGGALFSESP
LAPDAPEDEQEGTSGSGFLIDREGHILITNHYVHODATDIIRLHOMERXPASVY
TAPADLALRARSGARRPGAHAGGCRSPAGCRGGDAPFGLEFTVYTGIIIPAYK
RYIPMGVESIPONSOTDATALTPGNSGCPVLNRSCEVGVNTQILSPAGAVTGQNA
GVGFAPVNAVYKSLPLRLAGBELTVPRTIGVSNVLAQALTPSARALGLPQGLVYS
VEGTPAAAGRGGRSPQREPDGDIRLGGDIIYAVDQEVSTVODLQGLGKROSGD
EVLILRRDQGTILQKRLTS"
3113..3118
3127..3678
/note="OrfC: proline-rich protein"
/codon_start=1
/transl_table=11
/evidence="not_experimental"
/product="hypothetical protein"
/protein_id="AAF44053.1"
/db_xref="GI:7263005"
/translation="MTEHCAAKMSARSTRSGRCATVHSRSLGCVGESAAVREPAVY
ARLPVPFVVSAAPELGHLDSDMHRPAPSAALPVVFPVSVWNCSPAVPSRLP
RPSITPHIKISFCWGRGFSASPLILPKQITVPVGRAGGPPILRPHFMPARPCFERCP
AARPLVSLSCIVCAVYVLLSCI"
3691..4038
/note="OrfD"
/codon_start=1
/transl_table=11
/evidence="not_experimental"
/product="hypothetical protein"
/protein_id="AAF44054.1"
/db_xref="GI:7263006"
/translation="MTTAPLFSASAVISSRPGSVNVRSGCCGAAMWNCRRRTAP
ALAIRPASAAPSTTAATVARKVYQKRGIRVQRRAPCGVLMKSGPPAPDLAAQ
WMAIVAYAGRG"
4090..4551
/note="OrfE"
/codon_start=1
/transl_table=11
/evidence="not_experimental"
/product="hypothetical protein"
/protein_id="AAF44055.1"
/db_xref="GI:7263007"
/translation="MAAPTARTHSPPMLPLAGLMSRVAVALTTPPLPPPCSAV
TPPPACACFPVLTPTPTSPVPGRRSPTPPSPCFPPRGRSAPARMRPCGWPSRAA
CPGFSLAPLRPPPARGRRPPLQAPFAGSGRRRRCPACSSFPARSLSFL"
4666..4670
4678..5364
/note="OrfF"
/codon_start=1
/transl_table=11
/evidence="not_experimental"
/product="hypothetical protein"
/protein_id="AAF44056.1"
/db_xref="GI:7263008"
/translation="MSELITFRISPDHARALVALGADHTAPALARAADLVMOQISGP
APVTDAKRSAPARLRLPLGVYDGMDSAPLPARACKLNLTLTGAADPTGMAR
ALTGRDITAEQITCEGMOQLDIOHHRRAQIAAQQOQAPAPAPLITRTQARALHK
OLAGIVTEDERYTOAKIARRVESITLALNODEAALLVVAKEDEARBRAVVLNQD
PAAYGALLR"
5407..5973
5407..5412
/gene="ardu"
5422..5973
/gene="ardu"
5422..5973
/gene="ardu"
/codon_start=1
/transl_table=11
/product="Ardu"
/protein_id="AAF44049.1"
/db_xref="GI:7263001"

```

```

/translation="MTYTHPLIIEHRDPALMIGCI.AA1NACKLIGAMQASSDPAE
MGALEITLKASPEHPAEEMDMDNDMPTEAGTLLDLSAAYVALDALSRAPAEIV
AAWEMRGAEEMDADKLTDAYLGRFSDVEDYAAOYLDDSGALOEPEMLRPYINTAL
GRDMEINGVYEGKNGHFNHGA"
RBS      6148..6153
CDS      6164..6445
        /note="orfG"
        /codon_start=1
        /transl_table=11
        /evidence=not_experimental
        /product="hypothetical protein"
        /protein_id="AAFA4057.1"
        /db_xref="GI:7263009"
/translation="WQYARDMPKLPPEVAEDVLTVKLVLYTMANPGERSDRLMA
EIGVPGQAIKRLRGLVLYOEAPRGPKGYRAQTKAKKPADEHP"
RBS      6736..6741
CDS      6748..7749
        /note="orfH: alanine-rich protein"
        /codon_start=1
        /transl_table=11
        /evidence=not_experimental
        /product="hypothetical protein"
        /protein_id="AAFA4058.1"
        /db_xref="GI:7263010"
/translation="MQKLTAAALIFSAALATACGGGPMTPPOAPAPASADVDDPG
ARYASGTRKMSAQAVOSLGRDLATAPAPVYVYAVKMGTPGSTRYAGOKRAFLG
WKLSCAFPAHLTGATSGSYALAKTPDADVKTKTGLVSLPLASLNTSRSRT
SYAAVYGTCAVAASSALTLANGSAAEFGAGAPLVYCEAGVYATPTNPATLGLSD
MOSSTFTTSAGFAGLVAFQVDPKAVMLTKYQOAECEGADAAVAVASGWMOT
LSSPANTRRFPAVDEQKMWQEHYGLAVQFAKDCGTTHGETLLREGILCPGATPS"
gene     7953..9159
RBS      /gene="repU"
RBS      7953..7957
CDS      /gene="repU"
        7966..9159
        /function="Initiation of plasmid replication"
        /codon_start=1
        /transl_table=11
        /product="RepU"
        /protein_id="AAFA4050.1"
        /db_xref="GI:7263002"
/translation="MRHNPFWRLFAGLISQISLOEGNHAKPALTRISSASIS
PPVPPTIAOSQAPAPAPATATAPATAPLWTLACIDAPCRSARIPALFESALDVA
RACGAEHVTAVFHLPAELIMVHGLKSAFENQYIRRGVLAACAHMGDLGCS
VATGTLMAVTLKPRVLEKAGVRLMDHDMGKMRDLNADKAGTYINMLHPKS
VPBSQCEDEVLTGGLREKSAFPAVYEVVSTWALKSALPSDMNTYVPAPSLAEN
VWELADHRETTPRAYRAEIVDQARALAAAFGADSLGFMKRLIWNITRADACRDV
SDVGAIVLRVLRDVKHDQTAGCTPPRNLAAYVNAALADLLGLREYEGORGVAPDO
AQFAA"
protein_bind  complement(8717..8725)
        /note="putative"
        /bound_moiety="DnaA"
        complement(8745..8753)
        /note="putative"
        /bound_moiety="DnaA"
misc_feature  9164..9637
        /note="region of low %G+C; possibly involved in
        replication initiation"
RBS      10614..10619
gene     /gene="resU"
        10614..11590
        /gene="resU"
        10625..11590
        /gene="resU"
        /note="resolution of plasmid multimers - plasmid"

Query Match      64.0%; Score 16; DB 56; Length 11910;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
      8 catgccagcattcg 23
      |||||||||||||||

```

Db 11110 CATGCCAGCATTTCG 11095

RESULT 26
 DMBH48C10
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

REFERENCE

AUTHORS
 TITLE
 JOURNAL
 REMARK

REFERENCE

AUTHORS
 TITLE
 JOURNAL

COMMENT

Submitted (16-DEC-1999) European Drosophila Genome Sequencing Consortium
 Clone-BACH48C10: Config ID=1; Length=49587; Status=Finished
 Sequence submitted by Takis Benos, EMBL Outstation - The EBI, Hinxton, Cambridge, CB10 1SD, U.K.
 E-mail: benos@ebi.ac.uk on behalf of the European Drosophila Genome Sequencing Consortium. For further information see the European Drosophila Genome Sequencing Consortium's web site:
<http://edgp.ebi.ac.uk/>.

The syntax for the representation of annotation used in this record is documented at:
<ftp://ftp.ebi.ac.uk/pub/databases/edgp/sequence/annotation.README>
 Coding sequences are predicted from computer analysis, using both gene and CDS prediction programs and matches to other sequences. These predictions and matches have been evaluated by the annotators and may have been refined by hand (in which case a GeneFinder prediction will have no score. The annotators have also used their judgement on what matches to represent in this record. A far more complete annotation record is available from Flybase (<http://flybase.bio.indiana.edu/>) through the Flybase Annotation Object linked by the db_xref qualifier in the feature table. IMPORTANT: This sequence is NOT necessarily the entire insert of clone BACH48C10. It may be shorter, since we are minimising the overlap between clones to 100 bases, by trimming them. Clone BACH48C10 overlaps to the left with clone 82C7 The true left end of clone BACH48C10 is unknown
 Clone BACH48C10 overlaps to the right with clone BACH7M4 The true right end of clone BACH48C10 is unknown
 Sequence in absolute orientation with respect to chromosome.

FEATURES

source

gene
 CDS

1..49587
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone="BAC BACH48C10"
 complement(530..1394)
 /gene="EG:BACH48C10.1"
 complement(join(330..1009,1095..1394))
 /note="EG:BACH48C10.1"
 /gene="/prediction="(method:"genefinder",
 version:"084", score:"21.51")";
 prediction=(method:"gensecan", version:"1.0",
 score:"73.25")";
 /codon_start=1
 /protein_id="CAB63507.1"
 /db_xref="GI:6594143"
 /translation="MSKKYPNPYTKHFVSEKTRKSTCKRCHPDMAGRSENIMRLK
 RHHDPKSYLDEKREIRALAOQAKDANVPTKVLASKYPILYPTTKRQSTTDD
 DSQSHGEMTWTLNGEETGYDNKNNAVQFPGVFGKSEIDPEYQWNSANMELK
 ADVSVAAOTTNESSCYAFTSTTAKAKSPARAASAAALDQSPSDSDYFLRYLSKVLG

```
gene
2911..4557
/ gene="EG:BACH48C10.2"
/ join(2911..3312,3484..4557)
/ gene="EG:BACH48C10.2"
/ note="/prediction=(method:""genefinder"";
version:""084""; score:""52.65"");
/prediction=(method:""genscan""; version:""1.0"";
score:""112.75""; /match=(desc:""GH19593.5prime GH
Drosophila melanogaster head port2 Drosophila melanogaster
cDNA clone GH19593 5prime, mRNA sequence"";
species:""Drosophila melanogaster (fruit fly)"";
ranges:(query:2855..3324, target:EMBL:A1388635:1..470,
score:""2296.00""; (query:3478..3538,
target:EMBL:A1388635:453..513, score:""287.00"")),
method:""blastn""; version:""1.4.9""")
/codon_start=1
/protein_id="CAB63508.1"
/db_xref="GI:6594144"
/translation="MTVIRLTPRRITVDYCRRLSDVLTALRPFYQAIRTDGAIHP
AEKRVNESRILCSLESTIGPSASCLPDGLEYIISGSDSSSVSRILRVQV
DRESLDPRAVICOLISCKLPEKEDEQKRGQEMKEKGLKREPERERERVOA
PVOKNSALRELEFKDAKREHSEIIMRQALARSQASEVATLRGKIGIAALYQLGI
RGRYDGCMMERYHNCILKGLSDPKVAGOLGALRNRFVFAVYTGSLGDDVLM
ESGDSGSMGLAFSEMRLEHEQRLVLEKALSAVLGIRERRRVPDGCARAYV
GSLRVYRAVSEHLREDSVOELPASLODHIALIVPDSIPKYSQGLFLARSCGP
DLVEICRNLDVASGRPRRDIOYEROLMROCCQELGQRLSKDSVTTPDMWAI
RHLLSGVSCRSLSHITINMSPVDPDITVCIFWDMRSRK"
16590..19135
/ gene="EG:BACH48C10.3"
/ join(16590..17843..17908..18048,18117..18490,18556..19135)
/ gene="EG:BACH48C10.3"
/ note="the EST emb1A1063630A1063630 comes from the 5'
UTR"
/codon_start=1
/protein_id="CAB72239.1"
/db_xref="GI:6911910"
/translation="MSESTEGSDSDLYDPLAEELHNVOLYKHTRENIDALNAKEAN
LOEPMAVYLGESKALNTWVIGNPTPTSKLEIKPTVYISRLMEIPTEVOEL
TSLHLEAKQEIMERLNSODOEDSSVERPQPHYONOTOILOOOROLARVHHC
NDLTSLDGSPSGCGTLRQPKILRALPNOQRTSVEVISGRLCDMLMALKLO
LTPDMEVSTSGRHIIIPMHDTGLHNEELFVRLLDEPKLTHIKOTIKRTEFSL
VCEGCRRLITPGTSCOCNRRPHQRCANRPMLCQPPMDSYOLLAEENDNGVF
PGRTAVRPNMSSRSRSCSSSSSSSSSGNHRGPRPISDODKSNAPN
VCINTNRVTSVQRLIQARPPLPHCTDSNSTQASPTSLKHNRPRAADESN
KILLRLDAKSENNNIIAEELILGPRISSGFVYRAHMGPAVATLVANKTPSPA
QLQARNEVAAMLKTRHCNIIILPMGCVSPSLAIYOMCESSLKHYHSPTRKLN
TLIDIGROVAQMDYLHAKNIIHRLDKSNNIFLHEDISVKTIDREPLATAKTRMSEKO
ANOPGSIILMMAPEVIRMOELNPSYFQSVYVAFGVWELLAECLPGHISKDDILF
MYGRGILPRDMSQVSDAPQALKRLAEDCIKTPKDRPLFRLLMLNMLRTLPKIH
RSASEPNLTQSLONDEFLYLPSPKTPVNFNPFQFSAGNT"
complement(19754..20239)
/ gene="EG:BACH48C10.6"
/ complement(19754..20239)
/ gene="EG:BACH48C10.6"
/ note="/prediction=(method:""genefinder"";
version:""084""; score:""26.59"");
/prediction=(method:""genscan""; version:""1.0"";
score:""82.07""; /match=(desc:""SIMILARITY TO RIBOSOMAL
PROTEIN L14""; species:""Caenorhabditis elegans"";
ranges:(query:20027..20128, target:SPTREMBL:Q20430:36..3,
score:""94.00""; (query:19952..20011,
target:SPTREMBL:Q20430:67..48, score:""73.00"";
(query:19835..19924, target:SPTREMBL:Q20430:109..80,
score:""88.00"")), method:""blastx""; version:""1.4.9""")
/codon_start=1
/protein_id="CAB63504.1"
/db_xref="GI:6594140"
/translation="MALRIITKRVQAOPLAVTIGCOOTOLHTTTACEITRLARI
RVVNSDKKMAAGRRPCITIVYKNGVGTIDKVALIKGQMKGIIVGLKONOK
PKQKTFDSNNIIVLIDNDSPLGTRIHVPIPIRLTILKEKTLAKGADYTKVLAIASRY
V"
complement(21238..25103)
CDS
/ gene="EG:BACH48C10.5"
complement(join(21238..24581,24649..24951,25046..25103))
/ gene="EG:BACH48C10.5"
/ note="/prediction=(method:""genscan""; version:""1.0"";
score:""314.45""; /prediction=(method:""genefinder"";
version:""084""; /match=(desc:""GH02860.5prime GH
Drosophila melanogaster head port2 Drosophila melanogaster
cDNA clone GH02860 5prime, mRNA sequence"";
species:""Drosophila melanogaster (fruit fly)"";
ranges:(query:25045..25290,
target:EMBL:A1064644:369..124, score:""1221.00"";
(query:24786..24952, target:EMBL:A1064644:535..369,
score:""817.00""; method:""blastn""; version:""1.4.9"";
/match=(desc:""LD12603.5prime LD Drosophila melanogaster
embryo Bluescript Drosophila melanogaster cDNA clone
LD12603 5prime, mRNA sequence""; species:""Drosophila
melanogaster (fruit fly)""; ranges:(query:25040..25317,
target:EMBL:AA438345:305..28, score:""1363.00"";
(query:24647..24952, target:EMBL:AA438345:604..299,
score:""1422.00""; (query:24549..24579,
target:EMBL:AA438345:635..605, score:""155.00"")),
method:""blastn""; version:""1.4.9"";
/match=(desc:""bs12f02.y1 Drosophila melanogaster adult
testis library Drosophila melanogaster cDNA clone bs12f02
5' mRNA sequence""; species:""Drosophila melanogaster
(fruit fly)""; ranges:(query:23494..24011,
target:EMBL:A1945413:522..5, score:""2554.00""),
method:""blastn""; version:""1.4.9"";
/match=(desc:""GM04080.5prime GM Drosophila melanogaster
ovary Bluescript Drosophila melanogaster cDNA clone
GM04080 5prime, mRNA sequence""; species:""Drosophila
melanogaster (fruit fly)""; ranges:(query:22290..22894,
target:EMBL:AA802249:605..1, score:""2971.00"")),
method:""blastn""; version:""1.4.9"";
/match=(desc:""GH25306.5prime GH Drosophila melanogaster
head port2 Drosophila melanogaster cDNA clone GH25306
5prime, mRNA sequence""; species:""Drosophila melanogaster
(fruit fly)""; ranges:(query:21634..22176,
target:EMBL:A1405337:543..1, score:""2706.00"")),
method:""blastn""; version:""1.4.9""")
/codon_start=1
/protein_id="CAB72240.1"
/db_xref="GI:6911911"
/translation="MSSMKGSFATINWPRKKDKSTGSSMGKLSKSTQJLHLST
DQIDNDSQLEFNTITTPVPAATSNAPEDFRITVLIPKQILVARLGAAPVLSKLS
LVCDNKDLDDKLEFRSPVDASQVSTETIGVGSLEIRLCKRSSYDNFNDVHK
FQRTGVARDDLSKSSDSSSRHSKVTAKTPSPSSNSLSMDSTGMNTRTPVYVP
AKVLAAPPKRRKTRAPRPSQVCIPEGVDPDIPAAKSEPAVYAVIRPOLCSTPNL
SGPLEDCNGNSKSPDESDTGDHYATLDLKPAPAGSEPTPKRLQIKKAPAP
PDPOSGDNTSLASLPEPVPTNPINOPAPRITTPLPAPVAPVAYONGNCSGN
GNSKVLNRTPTPPEPRISGATEDDNDPQASQAGICEPARPSVSPPEPARKSQ
QESSDEPDAIKYNNRFLCOTSKKPPESATITAEALTRADILEEDNHHVDQASITN
GNLATPSESTSLTSMNLTDPISPPDSQDQWRNAGSLSPSSIDELVDLATIIN
QOARDLTKRQDPQWVEIESAKPNRLANSITTASTRIGADSFQAGASGGTEV
ASSGALSHRSSSHSVSLNKLEONGICIGQRSSSELSIESPLOSQEVYKILNKRK
NSLAEGSSASSTVTEQPKIKIOLIESIAKQSSSELSIESPLOSQEVYKILNKRK
SNPGVEKPTSPTRAVQOEPYSSPALOKSSPAVQORPTSPVEPSPVAVKVEBQAV
SLHOKELVYQKESPPLESKESPVYKSSPAVQORPTSPVEPSPVAVKVEBQAV
SITTPRTATPTPSPNALRPNNTMAETEOSQDEPAFSPKPLPTTVYSGPSPINFA
```

Query Match 64.0% Score 16; DB 6; Length 49587;

Best local Similarity 100.0%; Pred. No. 40;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 5 gtgatcgccagcatt 20

Db 48632 gtgatcgccagcatt 48647

RESULT 27 AC018267 60573 bp DNA HTG 09-DEC-1999
AC018267/c
LOCUS
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered

pieces.
 AC018267
 AC018267.1 GI:6552924
 VERSION
 KEYWORDS
 HTG: HTGS, PHASE2.
 SOURCE
 ORGANISM
 Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 60573)
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 Direct Submission
 Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
 Rockville, MD, USA
 This sequence was identified as CDM:10214125 by the submitter.
 For more information on this record e-mail to fly@celera.com.
 * NOTE: This is a 'working draft' sequence.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 Location/Qualifiers
 1..60573
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 BASE COUNT
 16964 a 14192 c 13649 g 15768 t
 ORIGIN
 Query Match 64.0%; Score 16; DB 65; Length 60573;
 Best Local Similarity 100.0%; Pred. No. 39;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 5 gtcacgcgcagcatt 20
 ||||||||||||||||
 Db 27002 GTCATCGCAGCATT 26987
 RESULT 28
 LOCUS
 ATAC009895 82289 bp DNA 24-JAN-2001
 DEFINITION
 Arabidopsis thaliana chromosome III BAC T21P5 genomic sequence,
 complete sequence.
 AC009895
 AC009895.4 GI:12408719
 VERSION
 KEYWORDS
 HTG.
 SOURCE
 ORGANISM
 Arabidopsis thaliana
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 Magnoliophyta; eudicotyledons; core eudicots; Kosidae; eurosids II;
 Brassicales; Brassicaceae; Arabidopsids.
 1 (bases 1 to 82289)
 Lin, X., Kaul, S., Town, C.D., Benito, M., Creasy, T.H., Haas, B.,
 Rongning, C.M., Koo, H., Fujii, C.Y., Ullrich, T.R., Barnstead, M.E.,
 Bowman, C.L., White, O., Niernman, W.C. and Fraser, C.M.
 Arabidopsis thaliana chromosome III BAC T21P5 genomic sequence
 Unpublished
 2 (bases 1 to 82289)
 Lin, X. and Kaul, S.
 Direct Submission
 Submitted (04-SEP-1999) The Institute for Genomic Research, 9712
 Medical Center Dr, Rockville, MD 20850, USA, xlin@tigr.org
 3 (bases 1 to 82289)
 Lin, X.
 Direct Submission
 Submitted (24-JAN-2001) The Institute for Genomic Research, 9712
 Medical Center Dr., Rockville, MD 20850, USA
 On Jan 24, 2001 this sequence version replaced gi:12280866.
 Address all correspondence to:
 Xiaoying Lin
 The Institute for Genomic Research
 9712 Medical Center Dr.
 Rockville, MD 20850, USA
 e-mail: xlin@tigr.org

BAC clone T21P5 is from Arabidopsis chromosome III and is near the
 molecular marker m1172.
 The orientation of the sequence is from SP6 to T7 end of the BAC
 clone.
 Genes were identified by a combination of three methods: Gene
 prediction programs including GRAIL (available by anonymous ftp
 from arthur.epm.ornl.gov), Genefinder (Phil Green, University of
 Washington), Genscan (Chris Burge,
 http://genome.stanford.edu/~chris/GENSCANW.html), and NetPlantGene
 (http://www.cds.dtu.dk/netgene/cosnetgene.html), searches of the
 complete sequence against a peptide database and the Arabidopsis
 EST database at TIGR (http://www.tigr.org/tdb/at.html).
 Annotated genes are named to indicate the level of evidence for
 their annotation. Genes with similarity to other proteins are named
 after the database hits. Genes without significant peptide
 similarity but with EST similarity are named as 'unknown' proteins.
 Genes without protein or EST similarity, that are predicted by more
 than two gene prediction programs over most of their length are
 annotated as 'hypothetical' proteins. Genes encoding tRNAs are
 predicted by tRNAscan-SE (Sean Eddy,
 http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats are
 identified by RepeatMasker (Arian Smit,
 http://ftp.genome.washington.edu/RM/RepeatMasker.html). Regions of
 genomic sequence that are not annotated as genes but have predicted
 exons by GRAIL are annotated as misc features.
 Location/Qualifiers
 1..82289
 /organism="Arabidopsis thaliana"
 /cultivar="Columbia"
 /db_xref="taxon:3702"
 /chromosome="III"
 /map="m1172"
 /clone="T21P5"
 1..24217
 /note="overlap with BAC clone T12J13
 (AC009327:65718..89934)."
 complement(join(8..215,451..587,675..857,959..1050,
 1459..1593,1682..1748,1837..2055,2146..2340,2492..2879,
 2979..3100))
 /gene="T21P5.1"
 complement(8..>3100)
 /gene="T21P5.1"
 /note="similar to hypothetical protein GB:CA838918
 [Arabidopsis thaliana]"
 complement(join(8..215,451..587,675..857,959..1050,
 1459..1593,1682..1748,1837..2055,2146..2340,2492..2879,
 2979..3100))
 /gene="T21P5.1"
 /note="unknown protein"
 /codon_start=1
 /protein_id="AA01600.1"
 /db_xref="GI:6017117"
 /translation="MSQISGNNNIPLESEYWSLVNKADEKSKINDLPYERAYEN
 YPFKEVYVYQMKFOENRQKLVKAGKMEIGELASRLQLYGHMYKTSAGYLS
 ESYVYFALITREYKEDGLFODLNIAKQKLFARFLAMVCLVGRBMYLADQFKR
 LIDCKRTROETPKEKVVAAQELVRLKSDTAPMNIPLRYSLVLDPNDACTPAS
 RSLRLDALISYYCNKVKSEITLDSFRMLQCLEMPGSLYQSGAKMGNAVGV
 ARINSQMDNPLPKNRAVLTSPSTHETLAVLATICELEPSHGILLIYLSKQIG
 QISSPLSARSATSVENILRDESHITKETSPSLQITPSGSSYIYPSDLVFTPKP
 LFIILDSSTVEKNICAGKGPALILSPSTPLISADFSPQSGSLFTFTSP
 VQAFCLSEISVSMEMENDIFTKAEKLISSMNMWASTLATSDILHVMQSLIDPPLR
 RLILRTFCRAVALIYVPPNNQNPQECPSLPESLTPYAPVQSAVFQMANVFGAT
 SKTITPDITMLSEF"
 join(3627..3661,3762..3843,3918..3995,4081..4140,
 4240..4344,4437..4547,4629..4670,4860..4973,5097..5254,
 5361..5494,5576..5682,5785..5827,5903..6144)
 /gene="T21P5.2"
 <3627..>6144
 /gene="T21P5.2"
 /note="predicted by genscan"
 join(3627..3661,3762..3843,3918..3995,4081..4140,
 4240..4344,4437..4547,4629..4670,4860..4973,5097..5254,
 5361..5494,5576..5682,5785..5827,5903..6144)

```

/gene="T21p5.2"
/note="hypothetical protein"
/codon_start=1
/protein_id="AA01580.1"
/db_xref="GI:6017097"
/translation="MDNRSSSESIKRHEIKDTJASRKLEDTNTKLIDPREMALVAK
VRSDEEHSILOERIAACLDKMLNKGELERKCADLRAVIDEKONSVTALNEL
ARKKQDLEENLKLAHDLKVEDEERYIEFTWSLIGLAEGVPRVANAATVSSGJLH
DOLWKTACANDRIKREISSIVENOPGTIESKDNHDPNSKTOASGSDNDYOTN
EOLLPMENTRNPNYHIMODESIRFENNOIGGSGOGIFPOPKRENGCYPLSSVAGKE
MIOEREEKAENSMMDAYNGNEEFASHYREGSPGIDGFOITIGDAIPGKVLGCGFVPR
GTTGCPQWVPHIEDGCTROYIEGATHEPYITVADVDKLAVECTPMDODGROVKYBD
ESGITSFNEYSKDVLLIMOGELVRLFANDONKIRCGNVPNGT"
/misc_feature
6225..6289
/note="exon predicted by xgrail, quality marginal"
6384..6503
/note="exon predicted by xgrail, quality excellent"
6794..6839
/note="exon predicted by xgrail, quality excellent"
6928..7033
/note="exon predicted by xgrail, quality marginal"
7293..7411
/rpt_family="(TAAA)n"
complement(7488..7547)
/note="exon predicted by xgrail, quality
excellent_shadowexon"
repeat_region
complement(7654..7784)
/rpt_family="(TAA)n"
complement(7819..7864)
/misc_feature
8054..8113
/note="exon predicted by xgrail, quality good_shadowexon"
8263..8380
/rpt_family="(TAAA)n"
complement(8592..8627)
/note="exon predicted by xgrail, quality
marginal_shadowexon"
marginal_shadowexon"
<8652..>9527
/gene="T21p5.3"
<8652..>9527
/gene="T21p5.3"
/note="contains zinc finger motif, C3HC4 type (RING
finger)"
8652..9527
/gene="T21p5.3"
/note="unknown protein"
/codon_start=1
/protein_id="AA01602.1"
/db_xref="GI:6017119"
/translation="MGSTGNPNMGTTYSYRDCSGVGVPCVMCYVIFPPPSFY
LDDEDDSSSDSPFLIALIGLASAFITVSYTLISKYCHRRHNSSTSAAIRI
VCSDFWGTNNNNNGATNPOTIGGGGGLDESILKSTVYKRRMGFPVSSSCS
VCLSEFOENESLRLPKCNHAFHPCIDTMLKSHSNCPICRAITVSSAVEIYDLTNO
OIVTENNISTGDSVYVYVNLDENSRNRRTYNEGSTPLXGDAGKRRRRRATIGVVK
LRRRCIDRRYMGD"
/misc_feature
complement(10838..10896)
/note="exon predicted by xgrail, quality
marginal_shadowexon"
11174..11254
/note="exon predicted by xgrail, quality good"
11690..11700
/note="exon predicted by xgrail, quality marginal"
join(<11814..12256,12381..12578,12676..13085,13162..13332,
13418..>13761)
/gene="T21p5.4"
<11814..>13761
/gene="T21p5.4"
/note="similar to unknown protein GB:AA032238 [Arabidopsis
thaliana]"
join(11814..12256,12381..12578,12676..13085,13162..13332,
13418..13761)
/gene="T21p5.4"
/note="unknown protein"

```

```

misc_feature
complement(14488..14563)
/note="exon predicted by xgrail, quality
excellent_shadowexon"
14893..14935
/note="exon predicted by xgrail, quality marginal"
join(<15141..15586,15675..15872,16258..16667,16780..16950,
17036..>17427)

Query Match
Best Local Similarity 100.0%; Score 16; DB 13; Length 82289;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 aggtgcacgcgcagca 18
Db 30302 AGGTGCATCGCCAGCA 30317
|||||

```

```

RESULT 29
LOCUS HS272E8/117968 bp DNA PRI 23-NOV-1999
DEFINITION Human DNA sequence from clone 272E8 on chromosome Xp22.13-22.31.
Contains a pseudogene similar to MDM2-Like P53-binding protein
gene. Contains STRs, GSSs and a CA repeat polymorphism, complete
sequence.
ACCESSION Z93929
VERSION Z93929.1 GI:3425887
KEYWORDS HTG; CA repeat polymorphism; MDM2-Like P53-binding protein.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 117968)
REFERENCE
AUTHORS Direct Submission
TITLE Submitted (17-AUG-1998) E-mail enquiries: humquery@sanger.ac.uk
JOURNAL Clone requests: clonerequest@sanger.ac.uk
COMMENT
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
IMPORTANT: This sequence is not the entire insert of clone 272E8.
It may be shorter because we only sequence overlapping sections
once, or longer because we arrange for a small overlap between
neighbouring submissions.
The true left end of clone 410B11 (286063) is at 117869 in this
sequence. The true right end of clone 390N22 (AL008711) is at 61633
in this sequence.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
This sequence was generated from part of bacterial clone contigs of
human chromosome X, constructed by the Sanger Centre Chromosome X
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/MGP/ChrX

```

272E8 is from the library RPEC11 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/VECTOR.pcrPAC2>.

FEATURES

source	location/Qualifiers
1.117968	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="X"
	/map="p22.13-22.31"
	/clone="RPI-272E8"
	/clone_11b="RPEC-1"
repeat_region	1370..1566
	/note="MER20 repeat: matches 2..218 of consensus"
repeat_region	/note="MER20 repeat: matches 2..218 of consensus"
	/complement(3645..3946)
repeat_region	/note="MER33 repeat: matches 324..2 of consensus"
	/complement(3864..3994)
repeat_region	/note="MER3 repeat: matches 131..1 of consensus"
	4194..4555
repeat_region	/note="MLR1A1 repeat: matches 5..365 of consensus"
	5433..5531
repeat_region	/note="MIR repeat: matches 87..188 of consensus"
	/complement(6029..6111)
repeat_region	/note="MIR repeat: matches 188..103 of consensus"
	6205..6322
repeat_region	/note="MIR repeat: matches 106..256 of consensus"
	7754
unsure	8261..9521
misc_feature	/note="predicted CpG island"
repeat_region	/complement(9741..9813)
	/note="MIR2 repeat: matches 146..74 of consensus"
repeat_region	/complement(11533..11655)
	/note="MIR2 repeat: matches 145..74 of consensus"
repeat_region	/note="MIR2 repeat: matches 145..74 of consensus"
	/complement(18288..19413)
repeat_region	/note="L1MD2 repeat: matches 1088..5 of consensus"
	/complement(19266..19525)
repeat_region	/note="L1 repeat: matches 5390..5135 of consensus"
	/complement(19531..22106)
repeat_region	/note="L1 repeat: matches 4931..2303 of consensus"
	/complement(22115..22419)
repeat_region	/note="AluX repeat: matches 301..1 of consensus"
	/complement(22424..22641)
repeat_region	/note="L1 repeat: matches 2314..2093 of consensus"
	/complement(23516..23641)
repeat_region	/note="MER25 repeat: matches 2036..1908 of consensus"
	25175..25396
misc_feature	/note="match: GSS U82600"
repeat_region	27948..28235
	/note="AluY repeat: matches 1..301 of consensus"
repeat_region	/complement(28665..29020)
	/note="THE1B repeat: matches 364..1 of consensus"
repeat_region	/complement(29889..30189)
	/note="AluJo repeat: matches 302..1 of consensus"
repeat_region	/complement(31338..31487)
	/note="MER5A repeat: matches 155..4 of consensus"
repeat_region	33121..33328
	/note="MIR repeat: matches 63..262 of consensus"
repeat_region	33424..33723
	/note="AluX repeat: matches 2..301 of consensus"
repeat_region	35811..36234
	/note="L1 repeat: matches 1575..2001 of consensus"
repeat_region	36723..39034
	/note="L1 repeat: matches 2121..4429 of consensus"
repeat_region	37893..38188
	/note="AluX repeat: matches 1..302 of consensus"
repeat_region	39040..39339
	/note="AluX repeat: matches 1..301 of consensus"
repeat_region	39353..40296
	/note="L1 repeat: matches 4449..5390 of consensus"
repeat_region	40148..41111
	/note="L1MD1 repeat: matches 1..968 of consensus"
repeat_region	/complement(41972..42157)
	/note="MLR1F repeat: matches 541..354 of consensus"
repeat_region	/complement(43606..43671)
repeat_region	/note="MIR2 repeat: matches 134..71 of consensus"
	45479..45585
repeat_region	/note="MER45 repeat: matches 26..133 of consensus"
	48242..48312
repeat_region	/note="MER45 repeat: matches 98..176 of consensus"
	49604..49826
repeat_region	/note="AluJb repeat: matches 82..301 of consensus"
	49936..50472
repeat_region	/note="L1PA13 repeat: matches 41..573 of consensus"
	/complement(50932..51134)
repeat_region	/note="MER20 repeat: matches 218..18 of consensus"
	51378..51596
repeat_region	/note="MER20 repeat: matches 1..215 of consensus"
	/complement(52293..52552)
repeat_region	/note="MLT1C repeat: matches 466..186 of consensus"
	/complement(52555..52908)
repeat_region	/note="THE1B repeat: matches 362..1 of consensus"
	/complement(52914..53071)
repeat_region	/note="MLT1C repeat: matches 182..8 of consensus"
	53212..53621
repeat_region	/note="MSMC repeat: matches 1..403 of consensus"
	55202..55295
repeat_region	/note="MIR2 repeat: matches 1..94 of consensus"
	55296..55504
repeat_region	/note="MER3 repeat: matches 1..209 of consensus"
	/complement(55783..56684)
repeat_region	/note="L1PA10 repeat: matches 911..1 of consensus"
	/complement(56534..57144)
repeat_region	/note="L1 repeat: matches 5390..4785 of consensus"
	/complement(57148..59003)
repeat_region	/note="L1 repeat: matches 4720..2850 of consensus"
	59735..59757
unsure	61714..61945
misc_feature	/note="match: SRS L24834"
repeat_region	/complement(63228..63358)
	/note="MIR repeat: matches 247..100 of consensus"
repeat_region	/complement(64335..64455)
	/note="L1ME3 repeat: matches 909..790 of consensus"
repeat_region	/complement(65455..65744)
	/note="AluJo repeat: matches 289..2 of consensus"
repeat_region	66343..66557
	/note="MIR repeat: matches 41..261 of consensus"
repeat_region	66912..67656
	/note="L1MC3 repeat: matches 1660..2409 of consensus"
repeat_region	/complement(68490..68794)
	/note="AluY repeat: matches 301..1 of consensus"
repeat_region	/complement(68873..69190)
	/note="AluSg repeat: matches 298..1 of consensus"
repeat_region	69217..69332
	/note="MER33 repeat: matches 1..324 of consensus"
repeat_region	/complement(69638..69703)
	/note="MIR repeat: matches 129..64 of consensus"
repeat_region	71210..71340
	/note="MIR repeat: matches 83..206 of consensus"
repeat_region	72426..72792
	/note="match: GSS A005396"
repeat_region	/complement(72923..72973)
	/note="MIR repeat: matches 139..89 of consensus"
repeat_region	/complement(75007..75196)
	/note="MER5A repeat: matches 189..1 of consensus"
repeat_region	/complement(75386..75425)
	/note="L1ME3 repeat: matches 406..367 of consensus"
repeat_region	75508..75800
	/note="AluJo repeat: matches 1..288 of consensus"
repeat_region	/complement(76744..76840)
	/note="MLR1F repeat: matches 182..91 of consensus"
repeat_region	/complement(78342..78521)
	/note="MIR repeat: matches 216..35 of consensus"
repeat_region	/complement(78968..79020)
	/note="MIR2 repeat: matches 144..112 of consensus"
repeat_region	/complement(79137..79186)
	/note="MIR2 repeat: matches 134..85 of consensus"
repeat_region	/complement(79151..79393)

```

repeatL_region /note="MIR repeat: matches 242. .3 of consensus"
                  complement(79492. .79666)
                  /note="MERB repeat: matches 178. .5 of consensus"
repeatL_region /note="MERB repeat: matches 178. .5 of consensus"
                  complement(79777. .79863)
repeatL_region /note="MERB repeat: matches 109. .27 of consensus"
                  complement(79894. .79970)
repeatL_region /note="MERB repeat: matches 11. .91 of consensus"
                  complement(80377. .80431)
repeatL_region /note="MIR repeat: matches 141. .87 of consensus"
                  complement(80389. .80462)
repeatL_region /note="MIR repeat: matches 252. .177 of consensus"
                  complement(80455. .80578)
repeatL_region /note="MIR repeat: matches 147. .29 of consensus"

Query Match      64.08; Score 16; DB 92; Length 117968;
Best Local Similarity 100.08; Pred. No. 37;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 atcgccagcattcga 24
    |||||
Db 97872 ATCGCCAGCATTTCGA 97857

RESULT 30
AP002817/c AP002817 140825 bp DNA PLN 12-AUG-2000
LOCUS      Oryza sativa genomic DNA, chromosome 1, PAC clone: P0699D11.
DEFINITION AP002817 BA000010
ACCESSION  AP002817.1 GI:9558510
VERSION     AP002817.1 GI:9558510
KEYWORDS
SOURCE      Oryza sativa (cultivar: Nipponbare) DNA, clone: P0699D11.
ORGANISM    Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
              Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae;
              Oryza.
REFERENCE   1 (bases 1 to 140825)
AUTHORS    Sasaki, T., Matsumoto, T. and Yamamoto, K.
TITLE      Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
           clone: P0699D11
JOURNAL     Published Only In Database (2000) In press
AUTHORS    Sasaki, T., Matsumoto, T. and Yamamoto, K.
TITLE      Submitted (26-JUL-2000) Takuji Sasaki, National Institute of
           Agricultural Resources, Rice Genome Research Program, Kannondai
           2-1-2, Tsukuba, Ibaraki 305-8602, Japan
           (E-mail: tsasaki@agr.affrc.go.jp, URL: http://rqp.dna.affrc.go.jp/,
           Tel: 81-298-38-7441, Fax: 81-298-38-7468)
COMMENT     The orientation of the sequence is from SP6 to T7 of the PAC clone.
           Genes were predicted from the integrated results of the
           following: GENSCAN1.0, BLASTN2.0, BLASTX2.0 as well as
           SplicePredictor (October 1998 version). The genomic sequence was
           searched against the non-redundant database NRP(PfR, SWISSPROT,
           GENEPIR, PDB) from MAF DNAbank and the cDNA sequence database at
           RGP. Protein similarities of the coding regions were searched
           against NRP with BLASTP2.0. ESTs represent the identified cDNA
           sequences using BLASTN2.0 with the corresponding DBJ accession no.
           and RGP clone ID.
           This sequence of P0699D11 clone has an overlap with P0462H08 clone,
           DBJ: AF002525 at the 5' end. The sequence of this clone starts at
           the position 114,739 of P0462H08. This sequence of P0699D11 clone
           has an overlap with P0469E09 clone, DBJ: AP001366 at the 3' end.
           The sequence of this clone ends at the position 59,396 of P0469E09.
           Detailed information on assemble quality together with annotation
           of this entry at
           http://rqp.dna.affrc.go.jp/Genomeseq.html.
           Location/Qualifiers
               1. 140825
                  /organism="Oryza sativa"
                  /cultivar="Nipponbare"
                  /db_xref="taxon:4530"
                  /chromosome="1"
                  /clone="P0699D11"

CDS
      join(527. .679, 925. 1114, 2076. .2467, 3067. .3555)
      /note="EST C22619(S11214) corresponds to a region of the
      Predicted gene.
      Similar to Arabidopsis thaliana chromosome 2, BAC clone
      117D12; putative protein kinase (AC006587)"
      /protein_id="BAB03429.1"
      /db_xref="GI:9558511"
      /translation="MWGIRRNNGHEATITFLRELADNTNNESTECCLRGSGFSVYK
      APLNDROYAVAKOLDINGLQGNREFLVEVLMSTLHHPNLVKGVCVGDGQRLIYE
      TPLGSLERLHDLRPGDEPLDWTTRMKIAADAAAGLSEYDEALPAVITSDIYSPG
      LLEGYNKATLSDPGKAKLGVGDKTHVTTRWGTHGKAPAEYLSGKLTISDIYSPG
      VVFLDLITGRALDSNRPDEQDLVAMARPLEKDDRKPPKADSDSLHGFPRGIFOA
      LAIAMLCOEKAKNRPSTREVALSVLASOFSQNTAAAHHTLPAGPVDVNDQIN
      QDLSLPSOHGVMPPLAGDHHVGVQVKNKCNSSSHRPRGAVTTPMGADREARALAEANY
      WYEARROKTKSKM"
      complement(6769. .9326)
      /note="ESTs AU093260(C52190), AU062991(C52190) correspond
      to a region of the predicted gene.
      hypothetical protein"
      /codon_start=1
      /protein_id="BAB03430.1"
      /db_xref="GI:9558512"
      /translation="WDFLEPTLEICRAFTIRGISDKITALSKPALVDASMSI
      AVDQSCVMDICDEFGNMGIFLDLVGSSDPHISDLASCLVLKAPREAVTGLTN
      LPLKLSVLDLKHGTCRLTRLLYCLAFSCQYLAQGMIVSISLSALRVEALVSARK
      GSHDCLADAASYGAELORLPRCG"
      complement(join(15587. .15727, 15868. .15915))
      /note="hypothetical protein"
      /codon_start=1
      /protein_id="BAB03431.1"
      /db_xref="GI:9558513"
      /translation="MNNFHYVEALIGRGHSTVYKGRKKSLFEFAVKSVDKSKVL
      NEVSAAPSSPCSFISSW"
      complement(join(18733. .18801, 19343. .19489, 19843. .19905,
      20016. .20087, 20611. .20691, 20769. .20819, 20976. .21072,
      21611. .21753))
      /note="ESTs C72327(E1437), AU078752(E1437) correspond to a
      region of the predicted gene.
      Similar to Archaeoglobus fulgidus competence-damage
      protein (AE000949)"
      /codon_start=1
      /protein_id="BAB03432.1"
      /db_xref="GI:9558514"
      /translation="WLSDGRIERAGRTKKNISSVNSGNTSTVEQITSKASIIYVG
      DEILFGEVDEKLAGLCKLHAIQWRSHVAVSNEDSYAEVEECKSTDDWFLVG
      GLGLHSDISLAVAKAFVYALAPDEFEYLSQLGDNYTGDRNEMALLPEGITEL
      HHKMLPLPLIKCKNVILATNVDLETETEMGCLLDQESGLVNAKSPVSKHICTSLD
      VKIAPVAKLCIDPSDYIGMT"
      join(30551. .30609, 30919. .31139, 31223. .31578)
      /note="Similar to Arabidopsis thaliana chromosome 4, BAC
      clone F10M23; hypothetical protein (AT035440)"
      /codon_start=1
      /protein_id="BAB03433.1"
      /db_xref="GI:9558515"
      /translation="MLGKEKRRRREGGENDIYAKKSMGPDPLPHRGISIKVP
      QIVIGTKNVLEPQNRKTAGSLNCTKPTRGSSVLTPLVLYVAESRMNPRHVKRRGH
      HGDIKQAGAAERDEEVYTDGLDHRQPHVKIPSEAAVYARRRMRMTQSSVGED
      GDAAVSPHKKVAVARRAAHSSVLEAGAGMKLKCDDLHNRNVLCTGTGLD"
      complement(join(31912. .32455, 32698. .33684, 34091. .34341))
      /note="hypothetical protein"
      /codon_start=1
      /protein_id="BAB03434.1"
      /db_xref="GI:9558516"
      /translation="MRKKRTNLLIVGEGATEFFFEERCRRRAGVSGSEVERRGDSKP
      SRGRHPATVARRAPLRROKARRDELDTASSGWTPESGWRPHRLPSPPLSLISH
      SLLIFSGRGEHCGGAASAGSGVSAAGSGAAGAAAGAAAGCGGCMGCAAPLSP
      RHPALSSAAGPARGDEDAGLVGSGEGTGGAGARARRRQIOGERRHKKRAPS
      VCCCFPTVAELVYLAVVCPALCHLAVRGGHRRVCAKOKRGELLTLDTVTPRS
      LAAAKARKVVEFPATPMABHILGAEVAKSKSLQALHVVAKLLSAGADARPA
      DLRYLLPVACPLSPYPIPLPRHRSLSCLPKRLSLTISIPKRGGRNDSSKHT
      TBNKTIKERENSTIIAQETCRSRSPAPMPTVARRAPLPNDSRVWGGAAPASAR
      EBGGGGGGDAALPACRLTAAYSLAASPKVTEPKRPVRLPJTNPINIRORAPQRCR

```

```

CDS
SSPSLSSSSLSGSSGKPARCGTARERLFWRRVRLQATLSIHAFAAHVIVLVV
VKLPARREACPRSTASAROGPPHRRR"
/join(35382..35472,36089..36172,36426..36530,37078..37337,
38995..39114)
/notes="hypothetical protein"
/codon_start=1
/protein_id="BAB03435.1"
/db_xref="GI:9558517"
/translation="MQLDASDFHGOELTLKLRDGLSSLESSGKSYDLVFPAAQOP
DAVFLPSPGEAAVGIARVSLVRYPPPELEKGLSLTPSSKRSAGSSKTRSR
FTSGSKRSSGSAOISLGOOSAPAHMKRKRDESLGHSNVGSKSAGSOVRGDS
GTSSEVPQTPVEKSKKKKKLDRPKRRTPYDRHKTQALDRTAAVSDSLYKSPKY
Q"
/join(40731..40753,43157..43538,45516..45821,47347..47893,
48657..48934,48983..50395)
/notes="Similar to Arabidopsis thaliana chromosome II BAC
clone T27A16: hypothetical protein (AC005496)"
/codon_start=1
/protein_id="BAB03436.1"
/db_xref="GI:9558518"
/translation="MYSLRPKIRPPFALAAADSLPMPPPPPGRPLPSWIRMRDR
RLLPYMSKSPGEGSQRGRENEEERERGERGEGGEGEELKGRGKEERE
EIRLGAGMRMRACGTSFADFLSRPNMSLSVLWLADEIKQLAARRRPIGIGG
HGAARDLPIDRSIVYAKLAGHSIQLASCSMLHLGPCSCAGLMRRRSASASL
LISSGMPNPTOPCKPNRSORPATLTLSLILSPPEFLSYELONHYSCIFISI
FSLNLCGRSPSPPTSTVPPPTAMRVRSIYIIPYDGVREMRKLAVYASMS
SLNPMEDQTVKICIGSSALESIGITPIHRAKPIAFGCFNPFPLKRYKSTALAH
AYAHSSSINKRAHPIRKILILYKQAPRLCSPKRNSSRIHNSAQCTHRLPAQ
RYVLDSEAHPPKQREERARETARPHDRSLSPNPPAPPTARVHSPVPSLP
DLRAHAYLLPRGHRHARVASALLASLRPLDRAAALVRRHVSLSAAARLGG
RGGGAAOASHLSLRAGHAADPHASASIVQASGVASARVDETAASADYVSM
NWKIDYVSGDLARARELEDMVGRNVVSMYVIGAVYAKOMQPEATITVFRMOVG
IIPDGYALLSVLSACGDLGVYDIGEVHVRVLRGICQELPLMNAITIDMYVCKSKK
ALEVFEGRMEKSLVTWTMTAGFALHGLGLEAVELFRMEKENVSNDTITLAVLSVQ
SHVGLDGLGRVYFKTWYQYKIKRVEHGCMLDLGRAGCLMEARGLIDMPFKMA
AIVGALLAARTHGTDELGEOLHLLIELEPHSGVYIILSNYAOEDMDVARELRI
SMRDGRIRNVPASISIDVGMVHEFTSRDGSHTLKHIREVLCAINSIKSVCHILL
PESLHYEEG"
complement(join(52441..53547,53887..54014,54128..54165,
54272..54487,54709..54870,55838..56016,56150..56256,
56375..56539,56658..56741,56821..57040))
/notes="Similar to Arabidopsis thaliana chromosome I BAC
clone T5A14: kinesin (AC005222)"
/codon_start=1
/protein_id="BAB03437.1"
/db_xref="GI:9558519"
/translation="MDVQPARTMRNLPTLTLMLGFKNLTPSMIESVHIIDGLSPT
KPMKVAVEKDENISDNTSEAKVOKIODELVLNAQLKQITLQREALNNYLDLG
NIRVFCIRPEHHEESYSRNLFPLDESANVFLVAETKRKQYFEDVFPDQSTQGVF
SEVBPVKSALDGYNCIFAYGOTSGKTYTMGKPTNLGVTIGTQITLONASGENE
RLFLTPSMLEIYMGNIHDLAPPSKINGIKNVSLSIKSPDGCITIEDIVATVANSF
QEVKRLTEMGTRLRSTASTMANSTSSRSCLINISLISLWATERKATSKLWMLDGG
SERLVKTKATGRKLRKGAKEKAINLSLALGDVIALQTKHPVPRNSKLTLLDLSGQV
ESTLMLVHISPDGDCETICTLGFATRSLRLESEEPKARKETLITLDLQKV
NDLEHCECIDERIKKLEESMEHITGQPTIYINPLKPPVRRRPSVYASVWPVNA
RRASSRLPRPKPTASSOHRIGINNPTIINRKPPVRRRPSVYASVWPVNA
APMOSSESSSECSMLTSDMNTSPSIRDTGCSODASYEIKQVIFSEHKESSDYTC
YTDPLAESKDIOIKTEKGIVIDNMVHOQVTEKSTPESKAWLIDPGVTEHIVS
SIPSPITMACTKEDSOVKDEVMGLTLOSSTYVEDIKOSTDNOFTAKELCTPPEKE
SSNNEVGHKHNHVVYHGRPPRSLOEBLEMLECTLEKPMNDSKSHRSHDKIKTGNVLSY
TGE"
/join(57603..57645,57742..57843,58195..58344,59425..59479,
59537..59589,61304..61785)
/notes="hypothetical protein"
/codon_start=1
/protein_id="BAB03438.1"
/db_xref="GI:9558520"
/translation="MKQDCKNNRCLPLQGSFSAVYIHOVYTINGECPTDPSRDSR
SEYAKIMIDYIPMHMATTPNGMRCLANENHSRAVIRKSGSPMCHAEILFSEATAIS
LIALSLAGKAFELSKRRSGDVHMLAISCVEAAAPLPLQPPILRPDDRRPAAWG
LCLSPASARRLSPKHOROPWRDRAAPAAAGRLSPNHRREIKRPLPISTSP
AIGRCRMRPLADATGELVELVSSKCRSCRVETIPSTVELLPMVVGELFLFSYVL
LPSPCSPAMSLAPSSRA"

```

CDS

complement(62330..63121)
/notes="Similar to Arabidopsis thaliana DNA chromosome 4,
BAC clone M3E9: hypothetical protein (AL022223)"
/codon_start=1
/protein_id="BAB03439.1"
/db_xref="GI:9558521"

Query Match 64.0%; Score 16; DB 13; Length 140825;
Best local Similarity 100.0%; Pred. No. 36;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 catcgccagcatcttg 23
DB 89622 CATCGCCAGCATTTTCG 89607

RESULT 31
AP001366/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AP001366 146081 bp DNA PLN 12-AUG-2000
Oryza sativa genomic DNA, chromosome 1, PAC clone: P0469E09.
AP001366 BA000010
AP001366.1 GI:7228436

Oryza sativa (cultivar: Nipponbare) DNA, clone: P0469E09.
Oryza sativa
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae;
Oryza.

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1 (bases 1 to 146081)
Sasaki, T., Matsunoto, T. and Yamamoto, K.
Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
clone: P0469E09
Published only in Database (2000) In press
2 (bases 1 to 146081)
Sasaki, T., Matsunoto, T. and Yamamoto, K.
Direct Submission
Submitted (09-MAR-2000) to the DDBJ/EMBL/GenBank databases. Takuji
Sasaki, National Institute of Agricultural Resources, Rice Genome
Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail: tsasaki@affrc.go.jp, URL: http://rtp.dna.affrc.go.jp/,
Tel: 81-298-38-7441, Fax: 81-298-38-7468)
The orientation of the sequence is from SP6 to T7 of the PAC clone.
Genes were predicted from the integrated results of the
following: GENSCAN 1.0, BLASTN 2.0, BLASTX 2.0 as well as
splice predictor (October 1998 version). The genomic sequence was
searched against the non-redundant database NRP (PIR, SWISSPROT,
GENPEPT, PDB) from MAF DNA bank and the cDNA sequence database at
RGP. Protein similarities of the coding regions were searched
against NRP with BLASTP 2.0. ESTs represent the identified cDNA
sequences using BLASTN 2.0 with the corresponding DDBJ accession
no. and RGP clone ID.

FEATURES
source
Detailed information on overlap and assemble quality together with
annotation of this entry at
http://www.dna.affrc.go.jp:82/genomicdata/GenomeFinished.html.

1. 146081
Location/Qualifiers
/organism="Oryza sativa"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/chromosome="1"
/clone="P0469E09"
/join(5902..5980,6229..6272,8134..8295)
/notes="hypothetical protein"
/codon_start=1
/protein_id="BAB03439.1"
/db_xref="GI:7228437"
/translation="MGAVRRSFLVWVFAILLTSMAGGRKMLINKHOVQSMETS
DDESMOGGEDEMLAMVHERILRQVKTNDYGTPTPTMAKRAKEIPN"
9673..9989

LTR
CDS
/notes="5' LTR"
/join(10528..12888,12673..13083,13175..13473,13550..13808,
13901..14083,14241..14306,14743..15045,15216..15479)
/notes="EST C28952(C62945) corresponds to a region of the

predicted gene.
Similar to maize transposon MUDR mudra protein isolog
(AC003981)"
/codon_start=1
/protein_id="BAA92398.1"
/db_xref="GI:7228438"
/translation="MTSKGIDRPAERFPOSISYMLNGFRIDQEVYTMSSVVSRT
EGYFWEIMPMSTAMRRYVMAERSPVITFVSQEDINVSQTEDEGPINAD
VVGFSQWENQOPREQAMGADGERGIIIVDMERDSDNEPADDASDEGDVW
AKTMANDESGVLIEGDHFWMEYKENEVIDGARDEKMEKVAIKHMAVSLQEEFV
KSTNYVYEVKMEDECPWYHAKGKNDWKYSYVTEHKKYLOGEYKHNITSEF
VASEYSSVGNNGEPEKSIIRHTEENKRYTISYKACRAKOKIEMRYGFEESYCN
LPRLATIAQNNNGEPEKSIIRHTEENKRYTISYKACRAKOKIEMRYGFEESYCN
TEMKYGKGOITAIIGDGNNOVYPAAFAPESENTSSWTFLEIRITAVRMKPNV
LIHRHAGMLKAITLQGWDEKRLPAKMPVRSWCMRHGANFYKQFNKHLMELE
KRLCAKNGEKEFENLMDKLDELTKQDEGRSPOVEDEPPIPLGALHDDPPMRER
SGSSIIRNFTQWIEENPEKESLIEDTDSRGYIMTNIAEYNNVMGVRVLPVAVI
EFLHTQAYFRDHYKKIGPSMADNNIVFGEQVRLPLHGAPANTLPYVSVGSKAM
LLOFTADLHPADNNSVRSLEAYITLFGWMTSTHGHAVDFRLVHYARSTYDQ
PODYQWMSGSAVIAATYRAICEACTMDAGITACCPMLQLMAERFAIGRVDS
APYGVSSAOMPDEGPTMGTYWCRGRRTAHVOYRGYPDEFEEDRLQSDVITWEY
TEEVAARAPLGLSLCTRODAWLTLLPWFDLFEVPHQFORVMQFORVPEGVAV
QPTVLPADHSLTRRGOLAGALMARVOOYVDWVLAEEVINEFPHTEENYRDYLEW
YLPRTRAVTFPPADPEPHVAAYTDAVPIHRDRDYFEGVADAARDISADITAVOCLR
GLHLDVEOARTFDRMOEKRAVMKRVSCSAVITPAPADLVLCRAPLASOCCIO
HRFTTYRPMCNFTLEWMDENPONDGRAYRSTRTSDYLRRDHECRITAAALEBH
VNLGLPTMRREPECRCGDCHYIRSVRHHTRGRCVCPNIVDDDFVLIHMMVEER
RVETASIDMRNRPGLFRKWSBRKCGCRDMQVATSPAEQTFRRRYFVCPNDDEVY
CLTIKIRYDHLKITQTE"

LTR
/note="3' LTR"
join(23418..23748,25558..25589,25715..26480,26838..26965)
/note="hypothetical protein"
/codon_start=1
/protein_id="BAA92399.1"
/db_xref="GI:7228439"
/translation="MGANCCITAKERPQCVTFLEVSFAFRNVRHSPSMFRMDNTHI
EDIMEMPALFSSNHSSTIRPETSSTIAPDGFSGNGSPDMNKILCHKSDKRSS
KIRSLDRGRSTSSNSPEAKLSKSLIDYVSAUSDMSISVSTPAISRADPSSS
RGSLPTDASMKARSPCYQLROYSDSKTISLRKLNFGASPEGSSSMVSCSN
DLASVSGHSGSSDSMSKRTSEMAVSOREWVSDELIGSVSKMTRSVASNNPTH
SPDEVCCLKLILKERSTWNAQOLAVAAVLGCHYHADCLSLTAEDAKPPCPV
CTHEKCTKILFGLLESKTKNKIPKNIYDVNLDGSKHKEKIPLEELYSQARASM
RELKSSRIGANDMTLNDLIRIEGS"
complement(join(27953..28624,29565..29918,30369..30666,
31026..31191))
/note="EST C96615(C10106),C26336(C12127),D21959(C10106)
correspond to a region of the predicted gene.
Similar to NAM (AL021889)"
/codon_start=1
/protein_id="BAA92400.1"
/db_xref="GI:7228440"
/translation="MESLRDMVLPGGFGHPKPKTELISHYKKKTKGOKLEVIIPDY
DIYKHPMDLAKCDVPTQDNKHFPAARBRKIPNSRSNRATVAGYMTSGDRAIK
MGOTCTKTKTLVFGHGRPTGRKRTIMHEYIDRECOQCPMDKDAVLCITRN
DWLPGNGELJNSDPHEPEYDAPSVISTEQMLPAEPVYGEAAVTAEPVPGCVTS
AITANIIPSPSDINLDDWLNELDFPDEQSLASDLSPDEQNVSSNGALAPVE
QDYSSPENNVDDTELDLEDVYNIILPGTDNDMONPLDPOYAYDWSGJOKE
ELMSPOANAPESQNSAADNGILIRYSKMTDPTSVOPFKSKQOAKMRGINKMASS
SESTNOTIKPENGSRIVEHOKNOADVASTKREDAKPTSELSSNGFLRGIRNAPAG
CSAPRMNMLIVAFAGVAVVAHICORIGLSORDOHT"
complement(36747..38998)
/note="hypothetical protein"
/codon_start=1
/protein_id="BAA92401.1"
/db_xref="GI:7228441"
/translation="MMGHPIYRADTARPPNNMAVPEPEYRHVDRHCTPEVPVLYLCRPD
DRGPCASLMAMAMGLIELVHGCCCTDDPEHGPSTI"
complement(39315..39726)
/note="3' LTR"
complement(join(40209..40396,40917..43545))
/note="EST C28952(C62945) corresponds to a region of the
predicted gene.

CDS
/note="EST C96615(C10106),C26336(C12127),D21959(C10106)
correspond to a region of the predicted gene.
Similar to NAM (AL021889)"
/codon_start=1
/protein_id="BAA92402.1"
/db_xref="GI:7228442"
/translation="MSSKVTFOIVHGEINIRFGPDGVDLSDFVMTSKGIDRPAERTFO
/SYSWLLRGERIDQEVYTMSSVVSRAIEGYFWEIMPMSTAMRRYVMAERSPVITFVSQEDINVSQTEDEGPINAD
VVGFSQWENQOPREQAMGADGERGIIIVDMERDSDNEPADDASDEGDVW
AKTMANDESGVLIEGDHFWMEYKENEVIDGARDEKMEKVAIKHMAVSLQEEFV
KSTNYVYEVKMEDECPWYHAKGKNDWKYSYVTEHKKYLOGEYKHNITSEF
VASEYSSVGNNGEPEKSIIRHTEENKRYTISYKACRAKOKIEMRYGFEESYCN
LPRLATIAQNNNGEPEKSIIRHTEENKRYTISYKACRAKOKIEMRYGFEESYCN
TEMKYGKGOITAIIGDGNNOVYPAAFAPESENTSSWTFLEIRITAVRMKPNV
LIHRHAGMLKAITLQGWDEKRLPAKMPVRSWCMRHGANFYKQFNKHLMELE
KRLCAKNGEKEFENLMDKLDELTKQDEGRSPOVEDEPPIPLGALHDDPPMRER
SGSSIIRNFTQWIEENPEKESLIEDTDSRGYIMTNIAEYNNVMGVRVLPVAVI
EFLHTQAYFRDHYKKIGPSMADNNIVFGEQVRLPLHGAPANTLPYVSVGSKAM
LLOFTADLHPADNNSVRSLEAYITLFGWMTSTHGHAVDFRLVHYARSTYDQ
PODYQWMSGSAVIAATYRAICEACTMDAGITACCPMLQLMAERFAIGRVDS
APYGVSSAOMPDEGPTMGTYWCRGRRTAHVOYRGYPDEFEEDRLQSDVITWEY
TEEVAARAPLGLSLCTRODAWLTLLPWFDLFEVPHQFORVMQFORVPEGVAV
QPTVLPADHSLTRRGOLAGALMARVOOYVDWVLAEEVINEFPHTEENYRDYLEW
YLPRTRAVTFPPADPEPHVAAYTDAVPIHRDRDYFEGVADAARDISADITAVOCLR
GLHLDVEOARTFDRMOEKRAVMKRVSCSAVITPAPADLVLCRAPLASOCCIO
HRFTTYRPMCNFTLEWMDENPONDGRAYRSTRTSDYLRRDHECRITAAALEBH
VNLGLPTMRREPECRCGDCHYIRSVRHHTRGRCVCPNIVDDDFVLIHMMVEER
RVETASIDMRNRPGLFRKWSBRKCGCRDMQVATSPAEQTFRRRYFVCPNDDEVY
CLTIKIRYDHLKITQTE"

LTR
/note="3' LTR"
join(23418..23748,25558..25589,25715..26480,26838..26965)
/note="hypothetical protein"
/codon_start=1
/protein_id="BAA92403.1"
/db_xref="GI:7228443"
/translation="MALPRGVRSPKPMMAAASASPPRPGHMLVLTGTVGRVSO
RLLAQWRRSGTCTSPAKTTELEMGMDASIFDRTSSRCW"
complement(join(48119..48257,48429..48703,49975..50139))
/note="hypothetical protein"
/codon_start=1
/protein_id="BAA92404.1"
/db_xref="GI:7228444"
/translation="MALVNDSDSEDDVNAAPCARAPVLYGAPGEVYLOLARLGE
ERDRARMAIISFCALSLRSMKTRITYVLCTLRNRYPLVTESKRTKCRIT
ASLASHGDELVCICRARLASSALDVRVEKILKTSDFVLMWLKRCIVINPVSIRH
GSSMYSTLPDAVPRSTERLAGVLDCTIGG"
complement(join(51918..52637,52684..53346))
/note="Similar to Arabidopsis thaliana chromosome 1 YAC
YUP8H12R sequence, C2-HC type zinc finger protein
Cε-MYTL (AC002986)"
/codon_start=1
/protein_id="BAA92405.1"
/db_xref="GI:7228445"
/translation="MASCINGELHVRVYHRLVKASDESIRPHVLAWSNLDLPRTI
OVSMEICYPKSTGDFHDVVAFAAGVPLLNHPPLAGITVNSPGSLPEICHNO
GAELVGEADVLAASIDYGTGASVYKTLIPYAGDVLAISYVSPAGGPTVANGTNN
VYVGSASLSMYSANSELARSGTIAGAARPHNDSDVRRSPSTYGADBEATPLUG
AROADIRLREOKARTVQAVSATLWMLAAVVGSRARRMWVWAGDERRLTLSS
BELRAMRSYVGNVTFTEFAVATEEIOKRLAIVASMAKALIAAPVGEHFDLSS
VEHRKAGORYIDTASVGLISPAVSATVAFEDVDTFGHMAHALPTSSSARLCT
GFVOIARPGGDSWIMASLMLPRLAALESDBRLGRIFKPYAVAYLGRPASGSSA
KRAGMITKTI"
complement(57570..58099)
/note="3' LTR"
complement(join(58018..58891,58980..62098))
/note="Similar to Arabidopsis thaliana chromosome II BAC
F9013 genomic sequence, putative retroelement pol
polyprotein. (AC006248)"
/codon_start=1
/protein_id="BAA92406.1"
/db_xref="GI:7228446"
/translation="MRVNLQAGLWTLAIDPQYAEFRDMAALSAILOAVREMLRGLA
KHDAKAAWDAIKTRVGVDRERAKQVAGRRGRRFREFRFEFARLTAIVYAD
IRDMGAWDEHVKKILRYRVPKKYKQVLAISLEOLDIKTMALAEIYGRSLTAVSD
DEGSDGKLYLTLEOMQARVYKORQEGSGKGGKGAATGQOHRKPGSGKGA
ATGANSRSDISRVKCFNCDERGHYARQCRPRORRORRANVQAADEEPILMAHYG
VSLGPAITLGTGPGGVHTEKVIILDHEDGGEETTRWFLDTGATNMTGVSAT
AELDTGVYGVTKFGDGSVIEIQGRGVVFRCKNGDRSLDTVYIIPILRNITISVGL

Similar to maize transposon MUDR mudra protein isolog.
(AC003981)"
/codon_start=1
/protein_id="BAA92402.1"
/db_xref="GI:7228442"
/translation="MSSKVTFOIVHGEINIRFGPDGVDLSDFVMTSKGIDRPAERTFO
/SYSWLLRGERIDQEVYTMSSVVSRAIEGYFWEIMPMSTAMRRYVMAERSPVITFVSQEDINVSQTEDEGPINAD
VVGFSQWENQOPREQAMGADGERGIIIVDMERDSDNEPADDASDEGDVW
AKTMANDESGVLIEGDHFWMEYKENEVIDGARDEKMEKVAIKHMAVSLQEEFV
KSTNYVYEVKMEDECPWYHAKGKNDWKYSYVTEHKKYLOGEYKHNITSEF
VASEYSSVGNNGEPEKSIIRHTEENKRYTISYKACRAKOKIEMRYGFEESYCN
LPRLATIAQNNNGEPEKSIIRHTEENKRYTISYKACRAKOKIEMRYGFEESYCN
TEMKYGKGOITAIIGDGNNOVYPAAFAPESENTSSWTFLEIRITAVRMKPNV
LIHRHAGMLKAITLQGWDEKRLPAKMPVRSWCMRHGANFYKQFNKHLMELE
KRLCAKNGEKEFENLMDKLDELTKQDEGRSPOVEDEPPIPLGALHDDPPMRER
SGSSIIRNFTQWIEENPEKESLIEDTDSRGYIMTNIAEYNNVMGVRVLPVAVI
EFLHTQAYFRDHYKKIGPSMADNNIVFGEQVRLPLHGAPANTLPYVSVGSKAM
LLOFTADLHPADNNSVRSLEAYITLFGWMTSTHGHAVDFRLVHYARSTYDQ
PODYQWMSGSAVIAATYRAICEACTMDAGITACCPMLQLMAERFAIGRVDS
APYGVSSAOMPDEGPTMGTYWCRGRRTAHVOYRGYPDEFEEDRLQSDVITWEY
TEEVAARAPLGLSLCTRODAWLTLLPWFDLFEVPHQFORVMQFORVPEGVAV
QPTVLPADHSLTRRGOLAGALMARVOOYVDWVLAEEVINEFPHTEENYRDYLEW
YLPRTRAVTFPPADPEPHVAAYTDAVPIHRDRDYFEGVADAARDISADITAVOCLR
GLHLDVEOARTFDRMOEKRAVMKRVSCSAVITPAPADLVLCRAPLASOCCIO
HRFTTYRPMCNFTLEWMDENPONDGRAYRSTRTSDYLRRDHECRITAAALEBH
VNLGLPTMRREPECRCGDCHYIRSVRHHTRGRCVCPNIVDDDFVLIHMMVEER
RVETASIDMRNRPGLFRKWSBRKCGCRDMQVATSPAEQTFRRRYFVCPNDDEVY
CLTIKIRYDHLKITQTE"

LTR
/note="3' LTR"
join(23418..23748,25558..25589,25715..26480,26838..26965)
/note="hypothetical protein"
/codon_start=1
/protein_id="BAA92403.1"
/db_xref="GI:7228443"
/translation="MALPRGVRSPKPMMAAASASPPRPGHMLVLTGTVGRVSO
RLLAQWRRSGTCTSPAKTTELEMGMDASIFDRTSSRCW"
complement(join(48119..48257,48429..48703,49975..50139))
/note="hypothetical protein"
/codon_start=1
/protein_id="BAA92404.1"
/db_xref="GI:7228444"
/translation="MALVNDSDSEDDVNAAPCARAPVLYGAPGEVYLOLARLGE
ERDRARMAIISFCALSLRSMKTRITYVLCTLRNRYPLVTESKRTKCRIT
ASLASHGDELVCICRARLASSALDVRVEKILKTSDFVLMWLKRCIVINPVSIRH
GSSMYSTLPDAVPRSTERLAGVLDCTIGG"
complement(join(51918..52637,52684..53346))
/note="Similar to Arabidopsis thaliana chromosome 1 YAC
YUP8H12R sequence, C2-HC type zinc finger protein
Cε-MYTL (AC002986)"
/codon_start=1
/protein_id="BAA92405.1"
/db_xref="GI:7228445"
/translation="MASCINGELHVRVYHRLVKASDESIRPHVLAWSNLDLPRTI
OVSMEICYPKSTGDFHDVVAFAAGVPLLNHPPLAGITVNSPGSLPEICHNO
GAELVGEADVLAASIDYGTGASVYKTLIPYAGDVLAISYVSPAGGPTVANGTNN
VYVGSASLSMYSANSELARSGTIAGAARPHNDSDVRRSPSTYGADBEATPLUG
AROADIRLREOKARTVQAVSATLWMLAAVVGSRARRMWVWAGDERRLTLSS
BELRAMRSYVGNVTFTEFAVATEEIOKRLAIVASMAKALIAAPVGEHFDLSS
VEHRKAGORYIDTASVGLISPAVSATVAFEDVDTFGHMAHALPTSSSARLCT
GFVOIARPGGDSWIMASLMLPRLAALESDBRLGRIFKPYAVAYLGRPASGSSA
KRAGMITKTI"
complement(57570..58099)
/note="3' LTR"
complement(join(58018..58891,58980..62098))
/note="Similar to Arabidopsis thaliana chromosome II BAC
F9013 genomic sequence, putative retroelement pol
polyprotein. (AC006248)"
/codon_start=1
/protein_id="BAA92406.1"
/db_xref="GI:7228446"
/translation="MRVNLQAGLWTLAIDPQYAEFRDMAALSAILOAVREMLRGLA
KHDAKAAWDAIKTRVGVDRERAKQVAGRRGRRFREFRFEFARLTAIVYAD
IRDMGAWDEHVKKILRYRVPKKYKQVLAISLEOLDIKTMALAEIYGRSLTAVSD
DEGSDGKLYLTLEOMQARVYKORQEGSGKGGKGAATGQOHRKPGSGKGA
ATGANSRSDISRVKCFNCDERGHYARQCRPRORRORRANVQAADEEPILMAHYG
VSLGPAITLGTGPGGVHTEKVIILDHEDGGEETTRWFLDTGATNMTGVSAT
AELDTGVYGVTKFGDGSVIEIQGRGVVFRCKNGDRSLDTVYIIPILRNITISVGL

Query Match	Similarity	100.0%	Score 16:	DB 13:	Length 146081:
Best Local	Similarity	100.0%	Pred. No. 36:		
Matches 16:	Conservative	0:	Mismatches	0:	Indels 0: Gaps 0:

QY	8	catgccagcatctgcg	23
Db	8193	CATGCCAGCATCTTCG	8178

RESULT	32
AC034114/c	AC034114

DEFINITION	AC034114	17-FEB-2001
Homo sapiens clone RP11-775E10, WORKING DRAFT SEQUENCE, 8 unordered pieces.	AC034114	
AC034114	GI:12958060	
HTG: HTGS_PHASE1: HTGS_DRAFT.		
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Enayratou, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.	
AUTHORS	1 (bases 1 to 154306)	
TITLE	2 (bases 1 to 154306)	
JOURNAL	Unpublished	
REFERENCE	2 (bases 1 to 154306)	
AUTHORS	Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F., Boguslavsky, L., Bonkhgalter, B., Brown, A., Burkett, G., Campopiano, A., Castle, A., Choe, P., Colangelo, M., Collins, S., Collumore, A., Cooke, P., DeRubeis, K., Dewar, K., Diaz, J.S., Dodge, S., Donnelly, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardina, S., Glendon, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Labèque, K., Lamazares, R., Landers, T., Lehotzky, J., Levine, R., Liu, C., Liu, G., Locke, K., MacDonald, P., Margulis, N., McCarthy, M., McKean, P., McKern, A., McKernan, K., McPheters, R., McFarlin, J., Menius, L., Milhova, T., Miranda, C., Mlenga, V., Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Olivar, T.M., Oliver, J., Peterson, K., Pierre, N., Plesniak, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Testa, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.	

TITLE	COMMENT
Submitted (04-APR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA	
On Feb 17, 2001 this sequence version replaced gi:8781909.	
All repeats were identified using RepeatMasker:	
Smit, A.F.A. & Green, P. (1996-1997)	
http://ftp.genome.washington.edu/BW/RepeatMasker.html	
Center: Whitehead Institute/ MIT Center for Genome Research	
Center code: WIBR	
Web site: http://www-seq.wi.mit.edu	
Contact: sequence_submissions@genome.wi.mit.edu	
Project Information	
Center project name: I4353	
Center clone name: 775_E.10	
Summary Statistics	
Sequencing vector: M13, M7815, 41% of reads	
Sequencing vector: Plasmid, n/a, 59% of reads	
Chemistry: Dye-terminator Big Dye, 100% of reads	
Assembly program: Phrap, version 0.960731	
Consensus quality: 151316 bases at least Q40	
Consensus quality: 152394 bases at least Q30	
Consensus quality: 153049 bases at least Q20	
Insert size: 151000: agarose-efp	
Insert size: 153606: sum-of-contigs	
Quality coverage: 10.6 in Q20 bases; agarose-efp	

```
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
```

```
1      41084: contig of 41084 bp in length
*      41085 41184: gap of          100 bp
*      41185 49589: contig of 8405 bp in length
*      49590 49689: gap of          100 bp
*      49690 58448: contig of 8759 bp in length
*      58449 58548: gap of          100 bp
*      58549 70162: contig of 11614 bp in length
*      70163 70262: gap of          100 bp
*      70263 85680: contig of 15418 bp in length
*      85681 85780: gap of          100 bp
*      85781 106804: contig of 21024 bp in length
*      106805 106904: gap of          100 bp
*      106905 139784: contig of 32880 bp in length
*      139785 139884: gap of          100 bp
*      139885 154306: contig of 14422 bp in length.
```

```
Location/Qualifiers
source
    1..154306
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone_id="RP11-775E10"
        /clone_lib="RPCI-11 Human Male BAC"
misc_feature
    1..41084
        /note="assembly_fragment"
        clone_end=SP6
        vector_side:left
    41185..49589
        /note="assembly_fragment"
    49690..58448
        /note="assembly_fragment"
    58549..70162
        /note="assembly_fragment"
    70263..85680
        /note="assembly_fragment"
    85781..106804
        /note="assembly_fragment"
    106905..139784
        /note="assembly_fragment"
    139785..154306
        /note="assembly_fragment"
        clone_end=T7
        vector_side:right"
BASE COUNT      43533 a   31887 c  31913 g   46271 t       702 others
ORIGIN
```

```
Query Match              64.0%: Score 16; DB 71; Length 154306;
Best Local Similarity 100.0%: Pred. No. 36;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps
```

```
OY      6 tgcacgcccagcattt 21
         |||
Db      78184 TGCAATGCCAGCATTTT 78169
```

```
RESULT  33
AC002457/c AC002457 177380 bp DNA PRI 21-DEC-1999
DEFINITION Homo sapiens BAC clone CTH-60P12 from Tq21, complete sequence.
ACCESSION AC002457
VERSION AC002457.1 GI:3947433
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

REFERENCE 1 (bases 1 to 177380)
 AUTHORS Smith,A., Beck,C. and Gibson,A.
 TITLE The sequence of Homo sapiens BAC clone CTB-60P12
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 177380)
 AUTHORS Waterston,R.
 TITLE Direct Submission
 JOURNAL Submitted (20-AUG-1997) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 REFERENCE 3 (bases 1 to 177380)
 AUTHORS Waterston,R.
 TITLE Direct Submission
 JOURNAL Submitted (02-DEC-1998) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 REFERENCE 4 (bases 1 to 177380)
 AUTHORS Waterston,R.
 TITLE Direct Submission
 JOURNAL Submitted (21-DEC-1999) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 COMMENT On Dec 2, 1998 this sequence version replaced gi:2337872.

----- Genome Center
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: http://genome.wustl.edu/gsc
 Contact: sapient@wustl.wustl.edu
 ----- Summary Statistics

 Center project name: H_RG060P12

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
 The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see
 http://www.nhgri.nih.gov/DIR/GRB/CHR7, send
 mailto:egreen@nhgri.nih.gov, or see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
 Clone CTB-60P12 is from the first release of the human BAC library CTB-978SK-B. The library contains cloned DNA from the male fibroblast cell line 978SK. See: Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); U-J. Kim et al., Genomics 34:213-8 (1996). This clone is available from Research Genetics, Inc. (http://www.resgen.com).
 VECTOR: pBelobAC11
 Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:
 The clone sequenced to the left is CTB-60N22, 200 bp overlap; the clone sequenced to the right is CTB-137N13, 200 bp overlap. Actual start of this clone is at base position 1 of CTB-60P12; actual end is at 177380 of CTB-60P12.

FEATURES
 source
 1..177380
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="7"
 /map="7q21"

gene	/clone="CTB-60P12"
	/clone_lib="CTB-978SK-B"
	complement(<131..112890)
	/gene="WUGSC:H_RG060P12.2"
	complement(join(<131..215,1834..1923,31059..31144,41160..41293,90777..90892,112769..112890))
	/gene="WUGSC:H_RG060P12.2"
	/note="Rap2 interacting protein; similar to U73941 (PDB:91916018); H_RG060P12.2"
	/protein_start=1
	/evidence=not_experimental
	/protein_id="AAC82532.1"
	/db_xref="GI:3947434"
	/translation="MASRSIGLGGTGGGGGKKSLSARAAVERRLITVCFESVKTLDSCPETIDSSPERNFAIIEQISRLKGQYTFEESPRESFQYTRACRKYSONSCISLNNENSSRAKRAIRVAMEKHUSEITSTALDKPFTTRRTPBDGATVGERANLACGLGLNADFSFLCKGGLDGSPPAVIDYPLKLYORY"
	480..791
repeat_region	/rpt_family="Alu"
repeat_region	4518..4824
repeat_region	/rpt_family="Alu"
repeat_region	4825..4937
repeat_region	/rpt_family="(TA)n"
repeat_region	5732..6004
repeat_region	/rpt_family="(TA)n"
repeat_region	6012..6047
repeat_region	/rpt_family="(L2)"
repeat_region	6143..6231
repeat_region	/rpt_family="(MIR)"
repeat_region	6232..7254
repeat_region	/rpt_family="(L2)"
repeat_region	7272..7412
repeat_region	/rpt_family="(L1)"
repeat_region	7420..7813
repeat_region	/rpt_family="(L1)"
repeat_region	7814..8150
repeat_region	/rpt_family="(MER1_type"
repeat_region	8151..8265
repeat_region	/rpt_family="(L1)"
repeat_region	8294..9183
repeat_region	/rpt_family="(Retroviral)"
repeat_region	9314..10441
repeat_region	/rpt_family="(L1)"
repeat_region	10437..12545
repeat_region	/rpt_family="(L1)"
repeat_region	12926..13352
repeat_region	/rpt_family="(MALR"
repeat_region	14327..14443
repeat_region	/rpt_family="(L1)"
repeat_region	14443..16612
repeat_region	/rpt_family="(L1)"
repeat_region	16719..16804
repeat_region	/rpt_family="(AT_rich"
repeat_region	17093..17116
repeat_region	/rpt_family="(AT_rich"
repeat_region	17492..17774
repeat_region	/rpt_family="(Alu"
repeat_region	18097..18135
repeat_region	/rpt_family="(AT_rich"
repeat_region	19255..19594
repeat_region	/rpt_family="(L1)"
repeat_region	19900..19946
repeat_region	/rpt_family="(GA)n"
repeat_region	19966..20245
repeat_region	/rpt_family="(L1)"
repeat_region	20549..20614
repeat_region	/rpt_family="(AT_rich"
repeat_region	20684..20816
repeat_region	/rpt_family="(AT_rich"
repeat_region	20817..21123
repeat_region	/rpt_family="(Alu"
repeat_region	21124..21134
repeat_region	/rpt_family="(AT_rich"


```

repeat_region      21136..21171
                    /rpl_family="(CA)n"
repeat_region      21478..21940
                    /rpl_family="L1"
repeat_region      22310..23003
                    /rpl_family="L1"
repeat_region      23135..23653
                    /rpl_family="MER21_g"
repeat_region      23658..24140
                    /rpl_family="L1"
repeat_region      24263..24304
                    /rpl_family="AT_rich"
repeat_region      24419..24447
                    /rpl_family="AT_rich"
repeat_region      24618..24865
                    /rpl_family="Alu"
repeat_region      25627..25650
                    /rpl_family="(TTAA)n"
repeat_region      25906..26152
                    /rpl_family="L2"
repeat_region      26153..26536
                    /rpl_family="AT_rich"
repeat_region      26654..27556
                    /rpl_family="Retroviral"
repeat_region      28842..28903
                    /rpl_family="(CATTA)n"
repeat_region      29888..30080
                    /rpl_family="MER1_type"
repeat_region      30382..30669
                    /rpl_family="Alu"
repeat_region      32790..32960
                    /rpl_family="MIR"
repeat_region      33062..33192
                    /rpl_family="L1"
repeat_region      33199..33998
                    /rpl_family="L1"
repeat_region      34000..34042
                    /rpl_family="(TAAA)n"
repeat_region      34043..34316
                    /rpl_family="Alu"
repeat_region      34414..34602
                    /rpl_family="L1"
repeat_region      35064..35263
                    /rpl_family="L1"
repeat_region      36255..36571
                    /rpl_family="Alu"

Query Match      64.0% Score 16; DB 85; Length 177380;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      3 aggtgcacgcacga 18
        |||
Db 48179 AGGTGCATGCCACGA 48164

RESULT 34
AC084136/c
LOCUS      AC084136      180700 bp      DNA
DEFINITION Homo sapiens clone RP13-506N4, WORKING DRAFT SEQUENCE, 32 unordered
            pieces.
ACCESSION AC084136
VERSION   AC084136.2 GI:12830241
KEYWORDS  HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE    human.
ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 180700)
            Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Beda,F., Boguslavsky,L.,
            Bouknight,B., Brown,A., Burkett,G., Campiano,A., Castle,A.,
            Chakrabarti,B., Collins,S., Collins,S., Collymore,A., Cooke,P.,
            Dearlano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P.,
            Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M.,
            Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L.,
            Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., Lacroque,K.,
            Lamazares,R., Landers,T., Lehotzky,J., Levine,R., Lien,C., Liu,G.,
            Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McKernan,K.,
            McPheters,R., Meldrim,J., Meneus,L., Minova,T., Menga,Y.,
            Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T.,
            O'Donnell,P., O'Neill,D., Olivari,T.M., Oliver,J., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Rieback,M., Riley,R.,
            Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P.,
            Sougniez,C., Spencer,B., Stange-Thoman,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
            Tirrell,A., Travers,M., Triggillo,J., Vassiliev,H., Viel,R., Vo,A.,
            Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
            Zimmer,A. and Zody,M.
            Direct Submission
            Submitted (13-OCT-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
            On Feb 14, 2001 this sequence version replaced gi:10800269.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html

            Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu

            Project Information
            Project name: 111406
            Center project name: 506_N_4
            Sequencing vector: Plasmid; n/a; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.960731
            Consensus quality: 167122 bases at least Q40
            Consensus quality: 176531 bases at least Q20
            Insert size: 183000; agarose-fp
            Insert size: 176000; sum-of-contigs
            Quality coverage: 3.7 in Q20 bases; agarose-fp
            Quality coverage: 3.8 in Q20 bases; sum-of-contigs

            NOTE: This is a 'working draft' sequence. It currently
            * consists of 32 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.

            1
            * 952 1051: gap of 100 bp
            * 1052 2064: contig of 1013 bp in length
            * 2065 2164: gap of 100 bp
            * 2165 3416: contig of 1252 bp in length
            * 3417 3516: gap of 100 bp
            * 3517 4816: contig of 1300 bp in length
            * 4817 4916: gap of 100 bp
            * 4917 6297: contig of 1381 bp in length
            * 6298 6397: gap of 100 bp
            * 6398 9144: contig of 2747 bp in length
            * 9145 9244: gap of 100 bp
            * 9245 11578: contig of 2334 bp in length
            * 11579 11678: gap of 100 bp
            * 11679 30905: contig of 19227 bp in length
            * 30906 31005: gap of 100 bp
            * 31006 32606: contig of 1601 bp in length
            * 32607 32706: gap of 100 bp

```

AUTHORS

TITLE
JOURNAL
COMMENT

Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Beda,F., Boguslavsky,L.,
Bouknight,B., Brown,A., Burkett,G., Campiano,A., Castle,A.,
Chakrabarti,B., Collins,S., Collins,S., Collymore,A., Cooke,P.,
Dearlano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P.,
Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M.,
Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L.,
Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., Lacroque,K.,
Lamazares,R., Landers,T., Lehotzky,J., Levine,R., Lien,C., Liu,G.,
Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McKernan,K.,
McPheters,R., Meldrim,J., Meneus,L., Minova,T., Menga,Y.,
Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T.,
O'Donnell,P., O'Neill,D., Olivari,T.M., Oliver,J., Peterson,K.,
Pierre,N., Pisani,C., Pollara,V., Raymond,C., Rieback,M., Riley,R.,
Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P.,
Sougniez,C., Spencer,B., Stange-Thoman,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
Tirrell,A., Travers,M., Triggillo,J., Vassiliev,H., Viel,R., Vo,A.,
Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
Zimmer,A. and Zody,M.

Direct Submission
Submitted (13-OCT-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Feb 14, 2001 this sequence version replaced gi:10800269.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu

Project Information
Project name: 111406
Center project name: 506_N_4
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 167122 bases at least Q40
Consensus quality: 176531 bases at least Q20
Insert size: 183000; agarose-fp
Insert size: 176000; sum-of-contigs
Quality coverage: 3.7 in Q20 bases; agarose-fp
Quality coverage: 3.8 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently
* consists of 32 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1
* 952 1051: gap of 100 bp
* 1052 2064: contig of 1013 bp in length
* 2065 2164: gap of 100 bp
* 2165 3416: contig of 1252 bp in length
* 3417 3516: gap of 100 bp
* 3517 4816: contig of 1300 bp in length
* 4817 4916: gap of 100 bp
* 4917 6297: contig of 1381 bp in length
* 6298 6397: gap of 100 bp
* 6398 9144: contig of 2747 bp in length
* 9145 9244: gap of 100 bp
* 9245 11578: contig of 2334 bp in length
* 11579 11678: gap of 100 bp
* 11679 30905: contig of 19227 bp in length
* 30906 31005: gap of 100 bp
* 31006 32606: contig of 1601 bp in length
* 32607 32706: gap of 100 bp

Phononayong, S., Pittman, G.S., Puri, V., Richards, S., Scheeler, F., Shapleton, M., Strong, R., Sytkas, R., Tector, C., Williams, S.M., Zaveri, J.S., Smith, H.O., Rubin, G.M. and Venter, J.C.
Sequencing of Drosophila chromosome 3R, region 88c-88c

TITLE
Unpublished
2 (bases 1 to 183439)

JOURNAL
Celinker, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G., Buttenhoff, C., Chame, M., Chavez, C., Chew, M., Ciesiolka, L., Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L., Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L., Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P., Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacled, J.M., Park, S., Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snij, E., Sytkas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and Rubin, G.M.

REFERENCE
Direct Submission
Submitted (14-JUN-1999) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Feb 24, 2001 this sequence version replaced g1:6623906.

COMMENT
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
This sequence was assembled using end sequences from a whole genome shotgun and from subclones of this BAC and its neighboring clones. For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive web site (<http://www.fruitfly.org/sequence/>) or send email to bdg@fruitfly.berkeley.edu.

FEATURES
source
1..183439
/organism="Drosophila melanogaster"
/strain="y: cn bw sp"
/db_xref="taxon:7227"
/chromosome="3R"
/map="88C-88C"
/clone="BACR45M03 (D718)"
/clone_11b="RPCT-98 (Roswell Park Cancer Institute Drosophila melanogaster BAC library, partial EcorI in pBACE3.6)"

BASE COUNT
53300 a 38849 c 38854 g 52436 t

ORIGIN

Query Match 64.0%; Score 16; DB 4; Length 183439;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 tcgcagcatttcgatt 25
|||||

Db 66707 TCGCAGCATTCGAT 66722

RESULT 36
AE003705 232744 bp DNA INV 05-OCT-2000
LOCUS Drosophila melanogaster genomic scaffold 142000013386035 section 30 of 105, complete sequence.
ACCESSION AE003705 AE002708
VERSION AE003705.1 GI:7299886
KEYWORDS HTG.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 232744)
Adams, M.D., Celinker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F., George, R.A., Lewis, S.E., Richards, S., Ashburner, M., Henderson, S.N., Sutton, G.G., Wortman, J.R., Yandell, M.D., Zhang, Q., Chen, L.X., Brandon, R.C., Rogers, J.H., Blake, R.G., Chame, M., Pfeiffer, B.D., Wan, K.H., Doyle, C., Baxter, E.G., Helt, G., Nelson, C.R., Gabor

Miklos, G.L., Abril, J.F., Agbayani, A., An, H.J., Andrews, P., Antkowiak, C., Baldwin, D., Bailey, R.M., Basu, A., Baxendale, J., Bayraktaroglu, L., Beasley, E.M., Beeson, K.Y., Benos, P.V., Berman, B.P., Bhandari, D., Bolshakov, S., Borrova, D., Botchan, M.R., Bouck, J., Brokstein, P., Brotler, P., Butts, K.C., Busan, D.A., Butler, H., Cadieu, E., Center, A., Chandra, I., Cherry, J.M., Cawley, S., Dahlke, C., Davenport, L.B., Davies, P., de Pablos, B., Delcher, A., Deng, Z., Meyers, A.D., Dew, I., Dietz, S.M., Dodson, K., Doup, L.E., Downes, M., Dugan-Rocha, S., Dunkov, B.C., Dunn, P., Durbin, K.J., Evangelista, C.C., Ferraz, C., Ferreira, S., Fleischmann, W., Foster, C., Gabrielian, A.E., Garg, N.S., Gelbart, W.M., Glasser, K., Glodek, A., Gong, F., Gorrell, J.H., Gu, Z., Guan, P., Harris, M., Harris, N.L., Harvey, D., Helman, T.J., Hernandez, J.R., Houck, J., Hostin, D., Houston, K.A., Howland, T.J., Wei, M.H., Idegawa, C., Jalali, M., Kalush, F., Karpen, G.H., Ke, Z., Kennison, J.A., Ketchum, K.A., Kimmel, B.E., Kodira, C.D., Kraft, C., Kravitz, S., Kuip, D., Lai, Z., Lasko, P., Lei, Y., Levitsky, A.A., Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Mattei, B., McIntosh, T.C., McLeod, M.P., McPherson, D., Merkulov, G., Milshina, N.V., Mobarry, C., Morris, J., Moshrefi, A., Mount, S.M., Moy, M., Murphy, B., Murphy, L., Muzny, D.M., Nelson, D.L., Nelson, D.R., Palazzolo, M., Pittman, G.S., Pan, S., Nusser, D.R., Pacled, J.M., Palazzolo, M., Reinert, K., Remington, K., Pollard, J., Puri, V., Reese, M.G., Reinert, K., Siden-Kiamos, I., Saunders, R.D., Scheeler, F., Shen, H., Shue, B.C., Sideris, K., Simpson, M., Skupski, M.P., Smith, T., Spier, E., Spradling, A.C., Stappleton, M., Strong, R., Sun, E., Sytkas, R., Tector, C., Turner, R., Venter, E., Wang, A.H., Wang, X., Wang, Z., Y., Weissman, D.A., Weinstein, G.M., Weissbach, J., Williams, S.M., Woodage, T., Wolley, K.C., Wu, D., Yang, S., Yao, Q., Zhao, Q., Zheng, L., Zheng, X.H., Zaveri, J.S., Zhan, M., Zhang, G., Zhou, X., Zhu, X., Smith, H.O., Zhong, F.N., Zhong, W., Zhou, X., Zhu, X., Smith, H.O., Gibbs, R.A., Myers, E.W., Rubin, G.M. and Venter, J.C.
The genome sequence of Drosophila melanogaster
Science 287 (5461), 2185-2195 (2000)

TITLE
JOURNAL
MEDLINE
2 (bases 1 to 232744)
20196006

REFERENCE
Adams, M.D., Celinker, S.E., Gibbs, R.A., Rubin, G.M. and Venter, C.J.
Direct Submission
Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA

FEATURES
source
1..232744
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="3R"
/complement(join(29274..29439,29524..30038,30107..>30980))
/gene="CG8464"
/product="CT24781"
/db_xref="FLYBASE:FBgn0008464"
/db_xref="FLYBASE:FBgn0038233"
/db_xref="FLYBASE:FBgn0038233"
complement(<29274..>30980)
/gene="CG8464"
/map="88C4-88C6"
/db_xref="FLYBASE:FBgn0008464"
/db_xref="FLYBASE:FBgn0038233"
complement(join(29274..29439,29524..30038,30107..30694))
/gene="CG8464"
/note="CG8464 gene product"
/codon_start=1
/db_xref="FLYBASE:FBgn0008464"
/db_xref="FLYBASE:FBgn0038233"
/protein_id="AAFP55062.1"
/db_xref="GI:7299887"
/translation="MALGSRRLVLEIFKRIASPIVSHANRRSSOLAIKEGPNNS
GNSGOYQNGEKGKRLVFEVPEFLGAVNSAIIQREDLPTLAASKMGRRRD
FNFIADVAGCSDSVVYEIKDTRHDFSCQPTVANSQFTIEQGLITNANHYI
NKHFTAVQVRLSDGRFPATIEDVDQTSDLATLRIVNNLSVMLGSSSTLRSENV
ALGSPALSNVTAGVISTORASOELIRNRDINYLQTDALITFGNSGFLVLDGE
AICGVNSLTAIGISFAIPIDYKVFLEAAERKGSAYKGYVVKRMTMLTLP
DILFEIKSSQNMPSNLTGCVLVMKVIYVSPHSGQLPGDVIHINKKEIKNSDVI
DALADNSKTLDIYILNGVKOMHVTITPDP"
join(<31212..31346,31407..31691,31752..>31922)
/gene="CG3351"

CDs

gene

mrna

mrna

```

/product="CT11255"
/db_xref="FLYBASE:FBan0003351"
/db_xref="FLYBASE:FBgn0038234"
<31212..>31922
/gene="CG3351"
/map="88C4-88C6"
/db_xref="FLYBASE:FBan0003351"
/db_xref="FLYBASE:FBgn0038234"
join(31212..31346,31407..31691,31752..31922)
/gene="CG3351"
/feature="CG3351 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0003351"
/db_xref="FLYBASE:FBgn0038234"
/protein_id="AAFS5063.1"
/db_xref="GI:7299888"
/translation="MSKAKGKLSKLTVERVTHRSKLTNTIPAGMAAGPILGMLG
ORAINIAPCKDINAKTAEMKEGVLPCKRISVNSDSYDLAIHHPATFPLKQAGTIO
RGMTPEKEVAGMTLKLHLEIAAIKIDPPVNLITMOOCMLISTARTCGIKVYRE
IDPAAYGEFLERKLIIEOQRRELCKRRAKMLRPG"
complement(join(32136..32708,32770..>32889))
/gene="CG8461"
/product="CT24741"
/db_xref="FLYBASE:FBan0008461"
/db_xref="FLYBASE:FBgn0038235"
complement(<32136..>32889)
/gene="CG8461"
/map="88C4-88C6"
/db_xref="FLYBASE:FBan0008461"
/db_xref="FLYBASE:FBgn0038235"
complement(join(32136..32708,32770..32835))
/gene="CG8461"
/feature="CG8461 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0008461"
/db_xref="FLYBASE:FBgn0038235"
/protein_id="AAFS5064.1"
/db_xref="GI:7299889"
/translation="WSSPSSSSASSEHSADNFTONAIREDLKGMSFEELMKLEEL
GAKVREAVLGSRSRPOKPRAKTDLRLNKNRPREMTTRQVPLGAHRVERKRV
RLDRPREKSGTVAETFEKKNQVDSOIRAKKVGOLKIDVEDDEAKRHIKDTMO
RLINKVEDKKKHTKOKOLKKEKRSATLKGOPHYLRKSTFLINDSEIKF"
complement(join(<34919..35540,35598..35757,35894..36047,
36334..36679,36886..37166,37233..37324,37381..37624,
37694..37981,38058..38138,42442..42707,42765..43010,
44497..44761,48465..>49118))
/gene="CG8279"
/product="CT24224"
/db_xref="FLYBASE:FBan0008279"
/db_xref="FLYBASE:FBgn0038237"
complement(<34919..>49118)
/gene="CG8279"
/map="88C5-88C6"
/db_xref="FLYBASE:FBan0008279"
/db_xref="FLYBASE:FBgn0038237"
complement(join(34919..35540,35598..35757,35894..36047,
36334..36679,36886..37166,37233..37324,37381..37624,
37694..37981,38058..38138,42442..42707,42765..43010,
44497..44761,48465..49118))
/gene="CG8279"
/feature="CG8279 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0008279"
/db_xref="FLYBASE:FBgn0038237"
/protein_id="AAFS5066.2"
/db_xref="GI:10726529"
/translation="MHGVPYRSSSSNMNTDVSPAGGAAPVEMSTSSSAAATNSAS
SSKRLNGANKTAISTAGGVTPGAVPGSGAIPASSSSGNOVLEHHHRSNNRP
AVNRSEETKLMTGSSSPSSOSPOLOASTOTISODRIAKASTTASODDVEVA
RLFEKPEAEKWLTERAPPEALSRQFTENRPHKRPSTVSDLPQMMASPTVVO
KSPRSLSNSASSLPECRHMLDLGSELFWELIDVANEJLDIVLCRKLIVANGLIT
HADRGSLFLAKGTPPTNKYLVAKIFDTQTKALDVTAKDVAISEEIIIPGIGIAGVAVQ
TKQINKIKAVYKQARPNCEIDLTGYKTNAIICMPCINVEGDIIGVAQIINKTNCME

```

```

FDEHDELPRRYLTFPGISITONMOTFEMSVQYRRNOIILNARSIFEBONNLECLVT
KIMTEAREILLRCRCSGYFLVDLDCASHLEKITEKPNOPATRAIKSADSEFEKMRN
RFLVLEFLGGEYQANVSRPSVSLSSFTLQIAPVATPGTVNICVIEVRHNO
IRADEIDSTQALICMKNMAOKRYGVNOLINKANGVFTSDASIFEAFIFCGIG
IHTOMYENACKILMAKOKVALIELCSYHAGASODPEKLELOVIAEASNYLSPFGLG
FELVDLDTCCRAVLRFMOCNLVSOFIIPDVLGRWLTVRKRYRKYKHNRHAIANA
OTEMFAMLTCKMERFMTDLEIIGLVACIChOLDHRCNNARQOTYESPLAILYTTST
TTLFPEPELDEPDDVYDSVYVSVLVLSVLSALTWATYKTERKT
RDNPMIALDSPEYRSVMKTVESAILISTDLAMVFKRRAFELVNGEFGDGEKKD
LTCGMATACDVSAIAKPWEVOHKAVALADEFPQGDLEKLOLTQVAMADREKRD
ELRKMVQYIDVLCPLRYVLCDTPEPMITPLEEGTLENSRNMODLAKEVGLMIDH
DTIDKEVEERFAACADEIIDIETVTLINCNOOSOGHSGDSSHPEHSGSRLSMKT
GALGKAVRSKLSKTLIYNSMDGSKPKTSKLTLESHVSEDMDKSPSPQPASSMKGR
MSASSSTSSGCGMVDSKRSKRLCALI"
join(40359..40547,40526..40826,40891..41076,41138..41210,
41284..41410,41461..42034,42102..>42262)
/gene="Cyp313a1"
/product="CT11271"
/db_xref="FLYBASE:FBan0003360"
/db_xref="FLYBASE:FBgn0038236"
<40359..>42262
/gene="Cyp313a1"
/feature="Cyp313a1"
/map="88C5-88C6"
/db_xref="FLYBASE:FBan0003360"
/db_xref="FLYBASE:FBgn0038236"
join(40391..40547,40526..40826,40891..41076,41138..41210,
41284..41410,41461..42034,42102..42262)
/gene="Cyp313a1"
/feature="Cyp313a1 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0003360"
/db_xref="FLYBASE:FBgn0038236"
/protein_id="AAFS5065.2"
/db_xref="GI:10726530"

Query Match
Best Local Similarity 64.0%; Score 16; DB 5; Length 232744;
Matches 16; Conservative 0; Pred. No. 34;
Mismatch 0; Indels 0; Gaps 0;

QY 10 tcgccagcatlccgat 25
Db 106438 TCGCCAGCATTTCCGAT 106453

RESULT 37
AC017740/c
LOCUS 237119 bp DNA HTG 10-DEC-1999
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
pieces
AC017740
VERSION AC017740.1 GI:6554259
KEYWORDS HTG; HTGS; PHASE2.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 237119)
AUTHORS Adams,M. and Venter,J.C.
TITLE Direct Submission
JOURNAL Submitted (10-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
COMMENT This sequence was identified as CDM:10211941 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
location/Qualifiers
1..237119
/organism="Drosophila melanogaster"

```

FEATURES
source


```

/protein_id="AA045778.1"
/db_xref="GI:7290318"
/translation="MPTMRHRRSSSSAAVEESRCRRGVGVGVDANKENGVHP
MSSPFGNALLALDLSNTHGKSPQRSFSGSPROATPOLAALRCIPTGGANAL
BDSQLSSRMGDTLLDRLDAIESARKEKRCIKTLPGACATTTTLADNMRSET
VHEMVEVPTPLKROVYRRKRNPKHTTTINOTJOSHOKKIEPIQLVPTKCIISFSS
SASDDEDOOVAKRSSIASPTTPSHCTTTTSSISSSSNGADMAQSOGSID
VSIADAKEOOLANVIRCRDLOSHSGSGMSINAVKVALSGAOPPGVGHSSGSGS
MSSGQRTAAVRHSHGRPTFDRENFQISSEBETAGOVLAWHMRDRHNSVPTGSE
WQAKYKSTERTTARERLAKLENAENAKRESEFGCSFPLNEIHPDSGSASYKHL
AOCAPPTSRHSQPKANAGODOKOETEAADODKODODSPAEMAAAVTYTPKPVATG
SGSGSNSAAMALNDEVISISIDSMKREDPLPOOPOMKLSKALHODADENFLR
FLIEDPPADGNASTTTOAOTSGSSASANSNNHINCGISGRSGRSMPSGGS
SVGAVROOGRCTPTMTKRLTREREGRFSTVWHPPEVEIEAGLSADRGILPG
DYVIFVKHNVYTPMEADVNLIRSQSSSLTLEIFRRSGAGATTITSTLGGNNVHS
TRICAAVGLSSEHTLATVATGTTVWSLORTTSTRLOPLANSMSRPATACGTTSSI
EAAKRRLHLPOVTFKESIVPVTDNRRRFLQLISREONFTALHFGVRFPOPIGR
KDLISPDHRTLFQNLDELRLAIEDLLEOLCSDDODOEPOMNFASRYVLSKTTTCA
AYKRCNGIKRADCVLNKSRGTSEFIATFEPAPRRRPLDLMTHRPHLOFREIL
KIMOLLAGNCHVDIEHKNSSTVINELQAAVREITYSSGLMEFLGGRPLTLQDLES
RMVTKCKPTLVAGROWTFCGDLSTRBSRVKPYWTLFSDIIVFAVSRDRLFI
TEEPPIANVVDSCFPHRKTTTEFLTVDPNGKLASPTGYCAPDLTRPKGARRS
LILRAPSLEIAWONLQKQIFLVNAALGSTPLSSPLSDPVLMTVPLVLSIGLTTA
SGSMKRLPSLDSIHLKQOQOQVRLFRSRKLDGSDYENLQLASAVRHHCSTTSS
ATQYQSHPPDPLCTTSLPSRTNSPARNRHNGOYGVGSGVNGGVSGFCGCTNNS
CLSSASITQTOVHTQTLNHSOTHSOTSTVLSISCKCLTVIPEVTSPPVQHA
SOKLEFSLSVGRSFLIDSGGVGGVGVSTLPTPTPTPTSTTQPCNSDAFSDVF
ATSTGKPMIPLAERFGSMDLLELDQLHEVNDLPSTDTDVEECIFLGDEDRGTE
GHGHDSDSDSDSDVVDPPGMLPTPTPTPSDLI"
complement(<3568..>54629)
/gene="EG:BACH7M4.4"
/product="CT33234"
/db_xref="FLYBASE:FBan0012496"
/db_xref="FLYBASE:FBan0040385"
/evidence=not_experimental
complement(<53568..>54629)
/gene="EG:BACH7M4.4"
/notes="CG12496"
/map="286-3A1"
/db_xref="FLYBASE:FBan0012496"
/db_xref="FLYBASE:FBan0040385"
/evidence=not_experimental
complement(53568..54629)
/gene="EG:BACH7M4.4"
/notes="EG:BACH7M4.4 gene product"
/codon_start=1
/db_xref="FLYBASE:FBan0012496"
/db_xref="FLYBASE:FBan0040385"
/evidence=not_experimental
/protein_id="AA045779.1"
/db_xref="GI:7290319"
/translation="MLKWTASQRLNIEPVHDDANOPORRRLORRORRATANK
ENVPPTVYTSPTVPTLRLNSPLVLAFLSDIRNVEPAVVRSAOPRVOSVYATL
PRFAEDYSPNGLPSGQHLALRGLAPLIDPFLQPRYIVGSGAVKILKROCKLDPAT
LEPELOEVAPEPTPKVEVRSPTPOPASTOMQDOTLRLIDAIILOSACANMSKK
PRSTFNLRRRTLVKQOMESNCPQEVNLSAPEDDASPLSGCLVPLPHNIMPTPE
PASPLAIPHLNLVOLATPVAPKSAPOCDNFCLTTPKALKRTRKEIVAKESPTRKYK
VDERNVEVGLALPSITYV"
complement(<70555..71849,71925..>71976))
/gene="gt"
/notes="Nucleotide sequence of the Celera sequence differs
from the published sequence for this transcript."
/product="CT23972"
/db_xref="FLYBASE:FBan0007952"
/db_xref="FLYBASE:FBan0001150"
complement(<70555..>71976)
/gene="gt"
/notes="CG7952"
/map="3A1-3A1"
/db_xref="FLYBASE:FBan0007952"

```

Query Match 64.0%; Score 16; DB 4; Length 289090;
 Best Local Similarity 100.0%; Pred. No. 34;

```

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 5 gtfcatcgccagcact 20
Db 13743 GTGCATCGCCAGCATT 13758
RESULT 39
NPTPLG/c
LOCUS NPTPLG
DEFINITION N. tabacum gene for pectate lyase.
ACCESSION X61102
VERSION X61102.1 GI:19981
KEYWORDS pectate lyase.
SOURCE common tobacco.
ORGANISM Nicotiana tabacum
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids
1; Solanales; Solanaceae; Nicotiana.
1 (bases 1 to 1389)
Lonsdale, D.M.
Submitted (25-JUL-1991) D.M. Lonsdale, Cambridge Laboratory, Centre
for Plant Science Research, Colney Lane, Norwich, Norfolk NR4 7UJ,
UK
2 (bases 1 to 1389)
Rogers, H.J., Harvey, A. and Lonsdale, D.M.
Isolation and characterization of a tobacco gene with homology to
pectate lyase which is specifically expressed during
microsporogenesis
Plant Mol. Biol. 20 (3), 493-502 (1992)
See X61101 for N. tabacum pectate lyase mRNA (partial).
LOCATION/Qualifiers
1..1389
/organism="Nicotiana tabacum"
/strain="Samsun"
/db_xref="taxon:4097"
/dev_stage="mature pollen"
/clone_lib="lambda-EMBL3"
/clone="G10"
<1..577
/number=1
join(1..577,676..976,1074..1389)
CDS
prim_transcript <1..>1389
join(1..577,676..976,1074..1389)
/codon_start=1
/product="pectate lyase"
/protein_id="CAA3414.1"
/db_xref="GI:19982"
/db_xref="SWISS-PROT:PA0972"
/translation="MDVYRIRISVFLVLLTPAALTATNIPRQLSNKKYKGPCRA
ENALDKCRCPNNAENKQKADALDEGSAIGKLGRLGYVVTDSDDVDVPPGRT
LRTGVLQEPMLITFGKNMFKLSRELIVTSNKTIDGRFNVIIONGAGIKIOSASNI
LISMRLHNIVPPGGLRSEDEHVGLSGDESDGISFSSHDIMWDHISMRATDGL
IDAAASNTITISNCHTDEKRWLFGANDHYVLDKMTITLAVNFEKRLDORPRC
RFGFHHVNDYTHMERYALGSSGATIIISQGRFLAEDLILYKTEYREKRTAYAE
MKMTWTISDDGDMENGAFTTPSDQNLKIDHLNLIKPEPSKVIITKFSGALSCV
KGRPC"
join(<1..577,676..976,1074..>1389)
578..675
/number=1
676..976
/number=2
977..1073
/number=2
1074..>1389
/number=3
BASE COUNT 427 a 270 c 305 g 387 t
ORIGIN

```

Query Match 60.0%; Score 15; DB 14; Length 1389;
 Best Local Similarity 100.0%; Pred. No. 2;le102;

ORGANISM
clone_lib:macaque brain cDNA library Qnpa clone:Qnpa-17839.
Macaca fascicularis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
Cercopithecinae; Macaca.

REFERENCE
AUTHORS
1 (sites), Hida, M., Kusuda, J., Tanuma, R., Iseki, K., Hirai, M.,
Terao, K., Suzuki, Y., Sugano, S. and Hashimoto, K.
Isolation of full-length cDNA clones from macaque brain cDNA
libraries
Unpublished (2000)

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
2 (bases 1 to 2616)
Hashimoto, K., Osada, N., Hida, M., Kusuda, J. and Sugano, S.
Direct Submission
Submitted (14-SEP-2000) to the DDBJ/EMBL/Genbank databases.
Katsuyuki Hashimoto, National Institute of Infectious Diseases,
Division of Genetic Resources; 23-1, Toyama 1-chome, Shinjuku-ku,
Tokyo 162-8640, Japan (E-mail: khashi@nih.go.jp,
url: http://www.nih.go.jp/yoken/genbank//,
tel: 81-3-5285-1111 (ex. 2120), fax: 81-3-5285-1181)
Lab host: TOP10
Vector: pME18S-FL3 (Acc.No. AB009864)
R. Site1: DraIII (CACTGTGTG)
R. Site2: DraIII (CAACATGTG)
Description: 1st strand cDNA was primed with an oligo(dT) primer
[ATGCGGCTTTTCTTTTCTTTT]; double-stranded cDNA was synthesized
using specific 5' and 3' primers and amplified by PCR. The PCR
product was digested with SfiI and size selection was performed to
exclude fragments <1.5kb. The SfiI-digested PCR product was cloned
into distinct DraIII sites of pME18S-FL3. XhoI sites just outside
were constructed by Sugano et al. (University of Tokyo, Institute of
Medical Science). Custom primer used for sequencing
{ 5' end primer [CGACCTGCACCTCGACGACA] };
{ 3' end primer [CTTTCGCTCTTAAAGCTGCG] }.

FEATURES
source
location/Qualifiers
1..2616
/organism="Macaca fascicularis"
/db_xref="taxon:9541"
/clone_lib="Qnpa-17839"
/clone_lib="macaque brain cDNA library Qnpa"
/dev_stage="adult"
/sex="male"
/tissue_type="brain parietal lobe"
BASE COUNT
739 a 527 c 572 g 778 t

Query Match
Best Local Similarity 60.0%; Score 15; DB 85; Length 2616;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 catgccacgacattc 22
|||||
Db 1743 CATGCCACGACATTTC 1729

RESULT 43
AF274670 3369 bp DNA PLN 13-AUG-2000
LOCUS Tetracentron sinense 26S ribosomal RNA gene, complete sequence.
DEFINITION
ACCESSION AF274670
VERSION AF274670.1 GI:9799468
KEYWORDS
SOURCE Tetracentron sinense.
ORGANISM Tetracentron sinense
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; Trochodendraceae; Tetracentroideae;
Tetracentron
1 (bases 1 to 3369)
Fishbein, M., Hirsch-Jetter, C., Soltis, D.E. and Hufford, L.
Phylogeny of Saxifragales (Angiosperms, Eudicots): Analysis of a
Rapid, Ancient Radiation

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
Unpublished
2 (bases 1 to 3369)
Fishbein, M., Hirsch-Jetter, C., Soltis, D.E. and Hufford, L.
Direct Submission
Submitted (30-MAY-2000) School of Biological Sciences, Washington
State University, Pullman, WA 99164-4236, USA
Location/Qualifiers
1..3369
/organism="Tetracentron sinense"
/specimen_voucher="Qiu 94166 (IND)"
/db_xref="taxon:13715"
1..3369
/product="26S ribosomal RNA"
BASE COUNT
806 a 824 c 1049 g 689 t 1 others

QY 3 aggtgcattgcacgc 17
|||||
Db 441 AGGTGCATGCCACG 427

Query Match
Best Local Similarity 60.0%; Score 15; DB 13; Length 3369;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 44
AF274671 3369 bp DNA PLN 13-AUG-2000
LOCUS Trochodendron aralioides 26S ribosomal RNA gene, complete sequence.
DEFINITION
ACCESSION AF274671
VERSION AF274671.1 GI:9799469
KEYWORDS
SOURCE Trochodendron aralioides.
ORGANISM Trochodendron aralioides.
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; Trochodendraceae; Trochodendroideae;
Trochodendron
1 (bases 1 to 3369)
Fishbein, M., Hirsch-Jetter, C., Soltis, D.E. and Hufford, L.
Phylogeny of Saxifragales (Angiosperms, Eudicots): Analysis of a
Rapid, Ancient Radiation
Unpublished
2 (bases 1 to 3369)
Fishbein, M., Hirsch-Jetter, C., Soltis, D.E. and Hufford, L.
Direct Submission
Submitted (30-MAY-2000) School of Biological Sciences, Washington
State University, Pullman, WA 99164-4236, USA
Location/Qualifiers
1..3369
/organism="Trochodendron aralioides"
/specimen_voucher="Qiu 94157 (IND)"
/db_xref="taxon:4407"
1..3369
/product="26S ribosomal RNA"
BASE COUNT
801 a 816 c 1060 g 690 t 2 others

QY 3 aggtgcattgcacgc 17
|||||
Db 441 AGGTGCATGCCACG 427

Query Match
Best Local Similarity 60.0%; Score 15; DB 13; Length 3369;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 45
PSE1AAMH 3677 bp DNA BCT 26-APR-1993
LOCUS PSE1AAMH/c
DEFINITION
P. syringae tryptophan 2-monooxygenase (1aam) and indoleacetamide
hydrolase (1aah) genes, complete cds.

ACCESSION M1035 M35690
 VERSION M1035.1 GI:151289
 KEYWORDS Indoleacetamide hydrolase; indoleacetic acid; tryptophan
 2-monooxygenase; tryptophan-2-oxyreductase.
 SOURCE *P. syringae* savastanoi (Strain EM2009) DNA, clones pUC2, pCP3, and pCP12.
 ORGANISM *Pseudomonas syringae*
 Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 Pseudomonas.
 REFERENCE 1 (bases 202 to 3677)
 YAMADA, T., PALM, C.J., BROOKS, B. and KOSUGE, T.
 Nucleotide sequences of the *Pseudomonas savastanoi* indoleacetic
 acid genes show homology with *Agrobacterium tumefaciens* T-DNA
 Proc. Natl. Acad. Sci. U.S.A. 82, 6522-6526 (1985)
 2 (bases 1 to 312)
 GAFFNEY, T.D., DA COSTA E SILVA, O., YAMADA, T. and KOSUGE, T.
 Indoleacetic acid operon of *Pseudomonas syringae* subsp. *savastanoi*:
 Transcription analysis and promoter identification
 J. Bacteriol. 172, 5593-5601 (1990)
 JOURNAL MEDLINE
 91008926
 COMMENT The *iaaH* and *iaaM* genes are part of an operon that is borne on a
 plasmid, pI_{IAA}, in *oleander* strains of the pathogen. There appears
 to be no genetic transformation of host tissue by *P. savastanoi*. No
Pribnow boxes were found, though potential ribosome binding sites
 are located at positions 435-439 and 2164-2168. Expression of *iaaH*
 depends upon the promoter for *iaaM*.
 Draft entry and computer-readable sequence for [1] kindly submitted
 by T. Yamada, 14-JAN-1986.
 Draft entry and computer-readable sequence for [J. Bacteriol.
 (1990) in press] kindly submitted
 by T.D. Gaffney, 28-JUN-1990.
 FEATURES
 source
 1..3677
 /organism="Pseudomonas syringae"
 /sub_species="savastanoi"
 /db_xref="taxon:317"
 19..47
 /note="Indoleacetic acid operon promoter"
 54..>3537
 /note="Indoleacetic acid mRNA"
 82..312
 /note="promoter proximal ORF"
 /codon_start=1
 /transl_table=11
 /protein_id="AAA25851.1"
 /db_xref="GI:151290"
 /translation="MPGRTPANWLDKAMVAVARIQRKPQAAVPAALLQAPHAMEFVW
 STTKFTTECHCPATPTPTQASAMASAVVSAV"
 454..2127
 /note="tryptophan 2-monooxygenase (EC 1.13.12.3)"
 /codon_start=1
 /transl_table=11
 /protein_id="AAA25852.1"
 /db_xref="GI:151291"
 /translation="MYDHFNSPSIDILYDGPFIKCKEPTGGISVSAGTPPRAVAV
 GAGISGLVAATELRAGVAKDVYVRESRDIGRVSQVDPQRRPRIAEMGAMRPPS
 ARLPHYIAKFGISTSTTPDPQVYDTELRKRRYHMPAGKRPPELRVRYEGWOST
 LSEGYLLEGGSLVAPLDITAMLSKGRLEEAALAWOGLNVPFDCSYNAIVCIFTGRH
 PGCGMARPEDELELGSIGSGFLPVOAGTEILRMVINGOSDRLIPDGISS
 LAARLADQSEFDGALDRVCFSVGRISREAEKTIITOTFAGORVDRVIYVTSNRM
 OMHCLTDESEFLSRDVARAVRETLTGSSKLPILTRKFWIKNNKLPITIOSDGLVYG
 VYCIQDPDEPEHGIVALLSTYMEDDAOKMLPCKKRCQVLYVDLIAIHPPTFASYL
 LPYDGYERYVLIHDMULTDPHSAQAFKINYPGEDVYSORLFPQPTANSFNKDTGLYL
 AGSCSFAGGMBGAVQTALNSACAVLRSTGOLSKGNPLDCINASYR"
 2170..3537
 /note="Indoleacetamide hydrolase"
 /codon_start=1
 /transl_table=11
 /protein_id="AAA25853.1"
 /db_xref="GI:151292"
 /translation="MHEITLESICQALADGETIAAELERRALDTEARLARLNCFTRE
 GNAVSGEGADHAMKCTPLMGKPVSFKNICVAGLPITFAGTGMGSGFVSDDDAIVSQ
 LRALGAVVAGKNNMHELSPGVTSINPHMCTVGNPVAFGYCGAGSSGSGSAAVAAGIAP

BASE COUNT 813 a 917 c 1046 g 901 t
 ORIGIN 1 bp upstream of EcoRI site.
 Query Match 60.0%; Score 15; DB 3; Length 3677;
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 7 gcatgcacgcatlt 21
 ||||||||||||
 DB 1729 GCATGCCGAGCATTT 1715

Search completed: October 9, 2001, 15:46:36
 Job time: 14782 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 15:52:25 ; Search time 469.84 Seconds
(without alignments)
33.410 Million cell updates/sec

Title: US-09-396-196f-6

Perfect score: 25
Sequence: 1 gccaggtgcacgcacacattcgtat 25

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 730101 seqs, 313950809 residues

Word size: 9

Total number of hits satisfying chosen parameters: 28031

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database:

N_Geneseq_0601:*

1: /SIDSL/gcgdata/geneseq/geneseq/NA1980.DAT:*

2: /SIDSL/gcgdata/geneseq/geneseq/NA1981.DAT:*

3: /SIDSL/gcgdata/geneseq/geneseq/NA1982.DAT:*

4: /SIDSL/gcgdata/geneseq/geneseq/NA1983.DAT:*

5: /SIDSL/gcgdata/geneseq/geneseq/NA1984.DAT:*

6: /SIDSL/gcgdata/geneseq/geneseq/NA1985.DAT:*

7: /SIDSL/gcgdata/geneseq/geneseq/NA1986.DAT:*

8: /SIDSL/gcgdata/geneseq/geneseq/NA1987.DAT:*

9: /SIDSL/gcgdata/geneseq/geneseq/NA1988.DAT:*

10: /SIDSL/gcgdata/geneseq/geneseq/NA1989.DAT:*

11: /SIDSL/gcgdata/geneseq/geneseq/NA1990.DAT:*

12: /SIDSL/gcgdata/geneseq/geneseq/NA1991.DAT:*

13: /SIDSL/gcgdata/geneseq/geneseq/NA1992.DAT:*

14: /SIDSL/gcgdata/geneseq/geneseq/NA1993.DAT:*

15: /SIDSL/gcgdata/geneseq/geneseq/NA1994.DAT:*

16: /SIDSL/gcgdata/geneseq/geneseq/NA1995.DAT:*

17: /SIDSL/gcgdata/geneseq/geneseq/NA1996.DAT:*

18: /SIDSL/gcgdata/geneseq/geneseq/NA1997.DAT:*

19: /SIDSL/gcgdata/geneseq/geneseq/NA1998.DAT:*

20: /SIDSL/gcgdata/geneseq/geneseq/NA1999.DAT:*

21: /SIDSL/gcgdata/geneseq/geneseq/NA2000.DAT:*

22: /SIDSL/gcgdata/geneseq/geneseq/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	25	100.0	1041	20 AAX01303	E. coli biotin syn
2	25	100.0	1084	10 AAN91329	E. coli Bto B gene.
3	25	100.0	1121	7 AAN60496	Sequence encoding
4	25	100.0	5872	15 AAO62386	Biotin-biosynthesi
5	16	64.0	1498	21 AAC51224	Arabidopsis thalia
6	16	64.0	1500	21 AAC36258	Arabidopsis thalia
7	15	60.0	2597	21 AAC93313	Human secreted pro
8	15	60.0	5773	21 AAC76014	Human ORFX ORF1569
9	14	56.0	549	21 AAC93854	Cat flea hindgut a
10	14	56.0	719	19 AAV20504	Conus geographus c
11	14	56.0	719	19 AAV17130	Conus geographus c

12	14	56.0	1190	21 AAC98831	Human pancreatic c
13	14	56.0	1344	21 AAZ56381	Escherichia coli f
14	14	56.0	2017	21 AAA69762	Human ovarian carc
15	14	56.0	2193	17 AAT30684	Kaposi's sarcoma a
16	14	56.0	2193	17 AAT16814	Kaposi's sarcoma a
17	14	56.0	2193	20 AAZ20914	Nucleotide sequenc
18	14	56.0	3593	21 AAZ39011	Mouse Ese2 coding r
19	14	56.0	4189	21 AAZ49334	Mouse Ese2 coding r
20	14	56.0	4313	14 AAQ38950	Mouse Ese2 full le
21	14	56.0	4625	21 AAZ39010	Mouse Ese2L coding
22	14	56.0	4975	21 AAZ39027	R. prowazekii S-la
23	14	56.0	5319	19 AAZ35235	Amycolatopsis medi
24	14	56.0	5676	19 AAV21186	Mouse Ese2L CDNA s
25	14	56.0	6014	21 AAZ39026	Kaposi's sarcoma a
26	14	56.0	20710	17 AAT30681	KSHV LTR DNA (nucl
27	14	56.0	20710	17 AAT16806	Amycolatopsis medi
28	14	56.0	35100	20 AAV73803	KSHV long unique c
29	14	56.0	53789	19 AAV21187	Oligonucleotide #3
30	14	56.0	137507	19 AAV19941	Human clone cq2838
31	13	52.0	22	19 AAV11150	Human secreted exp
32	13	52.0	51	21 AAZ76686	Human gene express
33	13	52.0	51	21 AAZ76687	Human gene express
34	13	52.0	286	21 AAA45502	Human gene express
35	13	52.0	300	20 AAZ14744	Plant microsatelli
36	13	52.0	300	20 AAZ14726	Hepatitis C virus
37	13	52.0	322	21 AAA31174	Secreted protein E
38	13	52.0	447	17 AAT7962	Fusarium venenatum
39	13	52.0	538	20 AAZ40788	Fusarium venenatum
40	13	52.0	629	21 AAF09259	Eucalyptus grandis
41	13	52.0	632	21 AAF09268	Human gene express
42	13	52.0	646	21 AAC55942	Human breast cancer
43	13	52.0	714	20 AAZ17005	Human gene express
44	13	52.0	725	22 AAF22436	Human gene express
45	13	52.0	749	20 AAZ16470	Human gene express

ALIGNMENTS

RESULT 1	
AAZ01303	standard: DNA; 1041 BP.
12-APR-1999	(first entry)
AAZ01303:	
XX	
AC	
XX	
DE	
XX	
XX	
XX	
KW	DAP aminotransferase; diamino-pelargonic acid; transgenic plant;
KW	biotin synthase; biotin production; vitamin H; Biot; ss.
XX	
OS	Escherichia coli.
XX	
PN	US5869719-A.
XX	
PD	09-FEB-1999.
XX	
PF	30-APR-1997; 97US-0846338.
XX	
PR	30-APR-1997; 97US-0846338.
PR	08-MAR-1995; 95US-0401068.
XX	
PA	(NOVS) NOVARTIS FINANCE CORP.
XX	
PI	Patton DA;
XX	
DR	WPI; 1999-152902/13.
DR	P-PSDB; AAW73906.
XX	
PT	Transgenic plants with high biotin levels - transformed with DNA
PT	encoding di-amino-pelargonic acid amino-transferase or biotin
PT	synthase

XX Example 2; Column 37-40; 34pp; English.

CC This sequence encodes the E. coli biotin synthetase (BioB). The gene can
 CC be used in the transgenic plant of the invention. The transgenic plant,
 CC plant cell or plant tissue is transformed with a chimeric gene encoding
 CC diaminopelargonic acid (DAP) aminotransferase or biotin synthase and
 CC produces more biotin than a non-transgenic plant, cell or tissue. The
 CC plant is used as an improved dietary source of biotin (vitamin H) for
 CC humans or animals.

XX SQ Sequence 1041 BP; 262 A; 273 C; 305 G; 201 T; 0 other;

Query Match 100.0%; Score 25; DB 20; Length 1041;
 Best Local Similarity 100.0%; Pred. No. 2.4e-05;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gcagggtgcatgcgcagcattcgat 25
 |||||
 DB 84 gcagggtgcatgcgcagcattcgat 108

RESULT 2

ID AAN91329 standard; DNA; 1084 BP.

AC AAN91329;

DT 15-FEB-1990 (first entry)

XX E.coli Bio B gene.

KW E.coli; Bio B gene; biotin.

OS Escherichia coli.

XX Key Location/Qualifiers

FT CDS 24..1064

PN GB2216530-A.

PD 11-OCT-1989.

PE 17-MAR-1989; 89GB-0006210.

PR 22-MAR-1988; 88GB-0006804.

XX 17-MAR-1989; 89GB-0006210.

PA (UKAG-) UK MIN. AGRIC. FISH.

PI Pearson BM, McKee RA;

DR WPI; 1989-295085/41. P-PSDB P91392

PT Plasmid contg. gene(s) for expression of biotin synthetase enzymes
 PT - derived from E.coli and capable of replication and expression in other
 PT microorganisms, esp. yeast.

XX Table 3; page 33-4; 52pp; English.

CC The gene can be used in a plasmid for expression of enzymes of the biotin
 CC synthetic pathway. Pre- control sequences for expression in S.cerevisiae
 CC are plasmids pMA51, pMA36c, pKV49 and pCK495, and plasmid pCK965 for
 CC Lactobacillus. Insertion of bio B improves biotin yields in
 CC microorganisms which export biotin, or enables growth in media contg.
 CC little or no biotin of organisms unable to synthesise biotin for their
 CC own use.

XX Sequence 1084 BP; 271 A; 286 C; 318 G; 209 T; 0 other;

Query Match 100.0%; Score 25; DB 10; Length 1084;
 Best Local Similarity 100.0%; Pred. No. 2.4e-05;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gcagggtgcatgcgcagcattcgat 25
 |||||
 DB 107 gcagggtgcatgcgcagcattcgat 131

RESULT 3

ID AAN60496 standard; DNA; 1121 BP.

AC AAN60496;

DT 17-OCT-1991 (first entry)

XX Sequence encoding biotin synthesising enzyme.

KW Biotin synthetic enzyme; E.coli; desthiobiotin; ds.

XX Key Location/Qualifiers

FT CDS 42..1082

PN JP61149091-A.

PD 07-JUL-1986.

PE 24-DEC-1984; 84JP-0272605.

PR 24-DEC-1984; 84JP-0272605.

XX (NIPS) NIPPON SODA KK.

PA WPI: 1986-216622/33.

DR P-PSDB; AAN60536.

XX Double stranded DNA encoding biotin synthesising enzyme -
 PT comprises transformed mutant E.coli strain contg. cyclic doubled
 PT stranded DNA encoding biotin synthetic biotin enzyme as a plasmid.

PS Disclosure; Page 534; 23pp; Japanese.

CC The sequence may be expressed by a transformed E.coli host, cultured
 CC in a medium containing desthiobiotin.

XX Sequence 1121 BP; 290 A; 301 C; 319 G; 211 T; 0 other;

Query Match 100.0%; Score 25; DB 7; Length 1121;
 Best Local Similarity 100.0%; Pred. No. 2.4e-05;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gcagggtgcatgcgcagcattcgat 25
 |||||
 DB 125 gcagggtgcatgcgcagcattcgat 149

RESULT 4

ID AAO62386 standard; DNA; 5872 BP.

AC AAO62386;

DT 16-NOV-1994 (first entry)

XX Biotin-biosynthesis genes contg. plasmid pB030A-15/9.

KW Biotin; expression; enterobacteria; vitamin H; synthesis;

KW plasmid; pB030A-15/9; bioB; bioF; bioC; bioD; bioA;

KW promoter; plac; biotin synthase; KAPA synthase;

KW 8-amino-7-oxononanoate synthase; pimeloyl-CoA; DTB synthase;

XX		dethiobiotin synthetase; DAPA synthase;
KW		S-adenosyl-L-methionine: 8-amino-7-oxononanoate aminotransferase;
KW		seborrhoea; dermatitis; ds.
XX		
OS		Escherichia coli DSM498.
XX		
FH	Key	Location/Qualifiers
FT	promoter	1..96
FT		/*tag= a
FT		/function= "promoter plac"
FT		/evidence= EXPERIMENTAL
FT		23..28
FT	-35_signal	/*tag= b
FT		/standard_name= "promoter plac"
FT		45..50
FT	-10_signal	/*tag= c
FT		/evidence= EXPERIMENTAL
FT		/standard_name= "promoter plac"
FT		105..109
FT	RBS	/*tag= d
FT		/evidence= EXPERIMENTAL
FT		/standard_name= "bioB RBS no. 9"
FT		117..1157
FT	CDS	/*tag= e
FT		/product= "biotin synthase"
FT		/evidence= EXPERIMENTAL
FT		/gene= "bioB"
FT		/number= 1
FT		1141..1146
FT	RBS	/*tag= f
FT		/standard_name= "bioF RBS"
FT		1154..2311
FT	CDS	/*tag= g
FT		/EC_number= 2.3.1.47
FT		/product= "KAPA synthase"
FT		/evidence= EXPERIMENTAL
FT		/gene= "bioF"
FT		/number= 2
FT		/standard_name= "8-amino-7-oxononanoate synthase"
FT		2284..2288
FT	RBS	/*tag= h
FT		/standard_name= "bioC RBS"
FT		2295..3050
FT	CDS	/*tag= i
FT		/function= "involved in pimeloyl-CoA synthesis"
FT		/product= "protein"
FT		/gene= "bioC"
FT		/number= 3
FT		3030..3033
FT	RBS	/*tag= j
FT		/standard_name= "bioD RBS"
FT		3043..3753
FT	CDS	/*tag= k
FT		/EC_number= 6.3.3.3
FT		/product= "DTB synthase"
FT		/evidence= EXPERIMENTAL
FT		/gene= "bioD15"
FT		/number= 4
FT		/standard_name= "dethiobiotin synthase"
FT		3712..3750
FT	misc_RNA	/*tag= l
FT		/note= "bioD15 substitution"
FT		3742..3746
FT	RBS	/*tag= m
FT		/standard_name= "bioA RBS"
FT		3750..5039
FT	CDS	/*tag= n
FT		/EC_number= 2.6.1.62
FT		/product= "DAPA synthase"
FT		/evidence= EXPERIMENTAL
FT		/gene= "bioA"
FT		/number= 5
FT		/standard_name= "S-adenosyl-L-methionine: 8-amino

```

FT      RBS      5088..5093      7-Oxononanoate aminotransferase"
FT      FT      /*tag= O
FT      FT      /standard_name= "ORF1 RBS"
FT      CDS      5098..5574
FT      FT      /*tag= P
FT      FT      /function= "unknown, involved in biotin synthesis"
FT      FT      /product= "protein"
FT      FT      /evidence= EXPERIMENTAL
FT      FT      /gene= "ORF1"
FT      FT      /number= 6
FT      FT      5583..5644
FT      FT      /*tag= q
FT      FT      /standard_name= "rho-independent transcriptional
FT      FT      stem_loop      terminator"
FT      FT      5583..5605
FT      FT      /*tag= r
FT      PN      W09408023-A.
FT      PD      14-APR-1994.
FT      XX      01-OCT-1993; 93WO-EP02688.
FT      XX      02-OCT-1992; 92CH-0003124.
FT      PR      15-JUL-1993; 93CH-0002134.
FT      XX      (LONZ ) LONZA AG.
FT      PA      Birch O, Brass J, Fuhrmann M, Shaw N;
FT      PI      WPI; 1994-135587/16.
FT      DR      P-PSDB; AAR51883, AAR51884, AAR51885, AAR51886, AAR51887, AAR63121.
FT      XX      Biotechnological biotin prodn. using enterobacterial biotin-gene
FT      PT      - providing vitamin H in high yield
FT      XX      Claim 1, Fig 6, Page 47-55 and 60-65; 92pp; German.
FT      PS      The sequence is derived from plasmid pB030A-15/9 confg. the
FT      CC      bioB, bioF, bioC, bioD and bioG genes responsible for biosynthesis
FT      CC      of biotin, arranged in a transcription unit. Microorganisms
FT      CC      confg. these DNA fragments or plasmids may be used in the prodn.
FT      CC      of biotin. Biotin (vitamin H) may prevent seborrhoea, dermatitis,
FT      CC      loss of appetite and tiredness.
FT      XX      Sequence 5872 BP; 1318 A; 1552 C; 1695 G; 1307 T; 0 other;
SQ
Query Match      100.0%; Score 25; DB 15; Length 5872;
Best Local Similarity 100.0%; Pred. NO. 2.3e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1 gcaagtcacgcgcagcattcgat 25
      ||||||||||||||||||||
Db      200 gcagtgcatcgccagcattcgat 224
RESULT 5
AAC51224
ID AAC51224 standard; DNA: 1498 BP.
XX
AC AAC51224;
XX
DT 18-OCT-2000 (first entry)
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 67727.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.

```

XX EPI033405-AZ.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123380.
PR 09-MAR-1999; 99US-0123348.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 07-MAY-1999; 99US-0132486.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134941.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135153.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137724.
PR 07-JUN-1999; 99US-0138094.
PR 08-JUN-1999; 99US-0138447.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139763.
PR 22-JUN-1999; 99US-0139817.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.

PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147302.
PR 09-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149962.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.

PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155586.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160778.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 64.0%; Score 16; DB 21; Length 1498;
Best Local Similarity 100.0%; Pred. No. 2; 6;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 agtgcacgcacga 18
|||||
Db 84 agtgcacgcacga 99

RESULT 6
AAC36258
ID AAC36258 standard; DNA: 1500 BP.
XX
AC AAC36258;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 13129.
XX
XX Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EPI033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX

PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126785.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129645.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 28-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.

```

PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145087.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145313.
PR 27-JUL-1999; 99US-0145318.
PR 28-JUL-1999; 99US-0145351.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 09-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 10-AUG-1999; 99US-0147935.
PR 11-AUG-1999; 99US-0148171.
PR 12-AUG-1999; 99US-0148319.
PR 13-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0148684.
PR 17-AUG-1999; 99US-0149368.
PR 18-AUG-1999; 99US-0149175.
PR 20-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 23-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 25-AUG-1999; 99US-0149930.
PR 26-AUG-1999; 99US-0150566.
PR 27-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151438.
PR 07-SEP-1999; 99US-0151930.
PR 10-SEP-1999; 99US-0152363.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155139.
PR 24-SEP-1999; 99US-0155486.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156556.

```

```

PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159337.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

Query Match 64.0%; Score 16; DB 21; Length 1500;
Best Local Similarity 100.0%; Pred. No. 2.0;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 3 aggtgcatgcgcagca 18
    |||
Db 84 aggtgcatgcgcagca 99

```

RESULT 7

```

AAC93313
ID AAC93313 standard; cDNA; 2597 BP.
XX
AC AAC93313;
XX

```

```

DT 16-FEB-2001 (first entry)
XX

```

```

DE Human secreted protein cDNA sequence #4.
XX

```

```

KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antileuc;
KW vulerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein; ss.
XX

```

```

OS Homo sapiens.
XX

```

```

PN W0200058495-A1.
XX

```

```

PD 05-OCT-2000.
XX

```

```

PF 23-MAR-2000; 2000MO-US07661.
XX

```

```

PR 26-MAR-1999; 99US-0126504.
PR 07-JAN-2000; 2000US-0174847.
XX

```

```

PA (HUMA-) HUMAN GENOME SCI INC.
XX

```



```

PI Rosen CA, Ruben SM, Komatsoulis G;
XX
XX WPI: 2000-611720/58.
DR P-PSDB: AAB51383.
XX
PT New nucleic acid molecules encoding 45 human secreted proteins for
PT diagnosing, preventing, treating or ameliorating medical conditions and
PT used as food additives or preservatives.
XX
XX Claim 1; Pages 327-328; 410pp; English.
XX
XX The invention relate to the isolation of genes AAC93310-C93354 encoding
CC 45 human secreted proteins AAB51380-B51423. The genes can be used to
CC generate fusion proteins by linking to the gene for the human
CC immunoglobulin G Fc portion (AAC93301) for increasing the stability of
CC the fusion protein as compared to the human protein only. The genes and
CC proteins are useful for preventing, ameliorating or treating medical
CC conditions, e.g. by protein or gene therapy. The genes are isolated
CC from a range of human tissues disclosed in the specification. The
CC nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer, and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, diabetes mellitus, Crohn's
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d)
CC wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
XX
XX Sequence 2597 BP; 721 A; 604 C; 534 G; 737 T; 1 other;
SQ
Query Match 60.0%; Score 15; DB 21; Length 2597;
Best Local Similarity 100.0%; Pred. No. 9.3;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 gcaagtgatcgcca 15
| | | | | | | | | | | | | | | | | |
DB 2256 gcaagtgatcgcca 2270

RESULT 8
AAC76014/C
ID AAC76014 standard; cDNA: 5773 BP.
XX
AC AAC76014;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human ORFX ORF1569 polynucleotide sequence SEQ ID NO:3137.
XX
XX Human: open reading frame; ORFX: detection: cytostatic; hepatotropic;
KW vulnervary; antiparasitic; antiparkinsonian; neutropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antineoplastic;
KW antiviral; antibacterial; antifungal; antihemetic; antithyroid;
KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antineoplastic disease; coagulation;
XX thrombosis; contraceptive; ss.
XX
XX Homo sapiens.
XX
XX WO200058473-A2.
XX
XX 05-OCT-2000.

```

```

XX
XX 31-MAR-2000; 2000MO-US08621.
PF
XX
XX 31-MAR-1999; 9905-0127607.
PR
XX
XX 02-APR-1999; 9905-0127636.
PR
XX
XX 05-APR-1999; 9905-0127728.
PR
XX
XX 30-MAR-2000; 2000US-0540763.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Shinkets RA, Leach M;
XX
XX WPI: 2000-602362/57.
XX
XX P-PSDB: AAB41805.
XX
XX Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease.
XX
XX Claim 5; Page 2354-2357; 5507pp; English.
XX
XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
XX which represent the human ORFX open reading frames 1 to 3161. The ORFX
XX sequences have activities such as: cytostatic; hepatotropic; vulnervary;
XX antiparasitic; antiparkinsonian; neutropic; neuroprotective;
XX osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
XX immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
XX antidiabetic; hypotensive; dermatological; immunosuppressive;
XX antineoplastic; antibacterial; antiviral; antifungal; antihemetic;
XX antithyroid; and antianemic. The sequences can be used for determining
XX the presence of or predisposition to, or preventing or treating
XX the pathological conditions associated with an ORFX-associated disorder. The
XX nucleic acids can be used to express ORFX proteins in gene therapy
XX vectors. The proteins and nucleic acids may be used to treat cancers,
XX proliferative disorders, neurodegenerative disorders, osteoarthritis,
XX graft vs host disease, cardiovascular disease, diabetes mellitus,
XX hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
XX erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
XX bacterial or fungal infection, malaria, autoimmune disorders, asthma,
XX allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
XX nocturnal haemoglobinuria, antineoplastic disease, to enhance
XX coagulation; to inhibit thrombosis; and as a contraceptive.
XX
XX Sequence 5773 BP; 1691 A; 1125 C; 1269 G; 1687 T; 1 other;
SQ
Query Match 60.0%; Score 15; DB 21; Length 5773;
Best Local Similarity 100.0%; Pred. No. 9.2;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 8 catcgacgatttc 22
| | | | | | | | | | | | | | | | | |
DB 4886 CATCGCACGATTTC 4872

RESULT 9
AAC93854/C
ID AAC93854 standard; cDNA: 549 BP.
XX
AC AAC93854;
XX
DT 19-FEB-2001 (first entry)
XX
DE Cat flea hindgut and Malpighian tubule (HMT) cDNA, SEQ ID NO:349.
XX
XX Cat flea; hindgut and Malpighian tubule nucleic acid; HMT;
KW flea infestation; vaccine; antiparasitic; therapeutic target;
KW diagnosis; detection; ss.
XX
XX Ctenocephalides felis.
XX
XX WO200061621-A2.
XX
XX

```

PD 19-OCT-2000.
 XX
 PF 07-APR-2000; 2000WO-US09437.
 XX
 PR 09-APR-1999; 99US-0128704.
 XX
 PA (HESK-) HESKA CORP.
 XX
 PI Brandt KS, Gaines PJ, Stinchcomb DT, Wisniewski N;
 XX
 DR WPI; 2000-656323/63.
 XX
 PR Flea Malpighian tubule and head and nerve cord tissue derived nucleic
 acids useful for the prevention, diagnosis and treatment of flea
 infestations -
 PR
 XX
 PS Claim 26; Page 361; 964pp; English.
 XX

The invention relates to novel cat flea (*Ctenocephalides felis*) nucleic acids which are expressed in hindgut and Malpighian tubule (HMT) tissue or head and nerve cord (HNC) tissue. The invention also relates to the encoded proteins. The invention additionally encompasses expression constructs, recombinant viruses and recombinant cells comprising the nucleic acids of the invention, recombinant production of the proteins, antibodies against the proteins, a method of identifying inhibitors of the proteins, and compositions comprising the inhibitors for administration to an animal. The nucleic acids, and the proteins they encode may be used in the prevention, treatment and diagnosis of diseases associated with flea infestations. For example, the nucleic acids may be used to produce an HMT or HNC protein according to standard recombinant DNA methodology by inserting the nucleic acids into a host cell and culturing the cell to express the protein. The HMT and HNC nucleic acids may also be used as DNA probes in diagnostic assays (e.g., PCR) to detect and quantitate the presence of cat flea or other homologous nucleic acid sequences in samples. They may also be used to study the expression and function of the proteins and their role in metabolism. The HMT and HNC proteins may be used as antigens in the production of specific antibodies, and in assays to identify modulators (agonists and antagonists) of HMT and/or HNC protein expression and activity. The anti-HMT/HNC protein antibodies and antagonists may also be used to downregulate protein expression and activity. The antibodies may also be used as diagnostic agents for detecting the presence of flea polypeptides in samples (e.g., by enzyme linked immunosorbent assay (ELISA)). The present sequence represents a cat flea HMT cDNA of the invention.

Sequence 549 BP; 155 A; 102 C; 130 G; 160 T; 2 other;

```
Query Match:      Score 14; DB 21; Length 549;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 14; Conservative 0; Mismatches 0; Gaps 0
```

Qy 8 caccgcagcattt 21
 |||
Db 421 CATCGCCAGCATTT 408

RESULT	10
AAV20504/c	
ID	AAV20504 standard; DNA; 719 BP.
XX	
AC	AAV20504;
XX	
DT	30-JUN-1998 (first entry)
XX	
XX	
DE	Conus geographus conantokin DNA.
XX	
XX	
KM	Conantokin: predatory cone snail; treatment; neurologic disorder;
KM	psychiatric disorder; anticonvulsant; neuroprotective; analgesic;
KM	HIV infection; ophthalmic indication; memory; Learning defect;
XX	cognitive defect; ss.
XX	
OS	Conus geographus..

XX	Key	Location/Qualifiers
FH	CDS	110., 412
PT		/+tag= a
FT		/product= conantokin
XX		
PN	W09803541-A1.	
XX		
PD	29-JAN-1998.	
XX		
PF	21-JUL-1997;	97WO-US12618.
XX		
PR	22-JUL-1996;	96US-0684742.
XX		
PA	(COGN-) COGNETIX INC.	
FA	(UTAH) UNIV UTAH RES FOUND.	
XX		
P1	Abogadie FC, Colledge C, Cruz LJ, Hillyard DR, Jimenez E;	
P1	Layer RT, McCabe RT, Olivera BM, Rivier JE, Shen GS;	
P1	Walker C, Zhou L;	
XX		
DR	WPI; 1998-120694/11.	
DR	P-PsDB; AAN48210.	
XX		
PT	New conantokin peptide(s) - useful for e.g. treating neurologic or	
PT	psychiatric disorders, or the management of pain	
XX		
XX	Claim 19; Pages 79-80; 122pp; English.	

The present sequence encodes Consu geographus conotoxin, peptide derivatives of which can be used to treat neurologic and psychiatric disorders, e.g. as an anticonvulsant, neuroprotective or analgesic agent. Neurologic and psychiatric disorders include epilepsy, convulsions, neurotoxic injury (associated with conditions of hypoxia, anoxia or ischaemia, which typically follow stroke, cerebrovascular accident, brain or spinal cord trauma, myocardial infarct, physical trauma, drowning, suffocation, perinatal asphyxia or hypoglycaemic events), neurodegeneration (associated with Alzheimer's disease, senile dementia, Amyotrophic Lateral Sclerosis, Multiple sclerosis, Parkinson's disease, Huntington's disease, Down's Syndrome, Korsakoff's disease, schizophrenia, AIDS dementia, multi-infarct dementia, Binswanger dementia and neuronal damage associated with uncontrolled seizures), chemical toxicity (such as addiction, and morphine, opiate, opioid and barbiturate tolerance), pain (acute, chronic, migraine), anxiety, major depression, manic-depressive illness, obsessive compulsive disorder, schizophrenia and mood disorders (such as bipolar disorder, unipolar depression, dysthymia and seasonal affective disorder) and dystonia (movement disorder) and sleep disorder, muscle relaxation and urinary incontinence. The peptide can also be used to treat HIV infection, ophthalmic indication and memory, learning or cognitive defects.

Query Match	56.0%;	Score 14;	DB 19;	Length 719;
Best Local Similarity	100.0%;	Prod. No. 35,		
Matches 14;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0
SQ Sequence 719 BP; 217 A; 171 C; 149 G; 182 T; 0 other;				

RESULT	11
AAV17130/C	
ID	AAV17130 standard; DNA; 719 BP
XX	
AC	AAV17130;
XX	
DT	30-JUN-1998 (first entry)
XX	

```

DE  Conus geographus conantokin DNA.
XX
XX  Conantokin; predatory cone snail; treatment: neurologic disorder;
KW  psychiatric disorder; anticonvulsant; neuroprotective; analgesic;
KW  HIV infection; ophthalmic indication; memory; learning defect;
KW  cognitive defect; ss.
XX
XX  Conus geographus.
OS
XX
XX  Key      Location/Qualifiers
FH  CDS      110..412
FT          /*tag= a
FT          /product= conantokin
XX
XX  WO9803189-A1.
PN
XX
XX  29-JAN-1998.
PD
XX
XX  21-JUL-1997; 97WO-US12652.
PF
XX  06-DEC-1996; 96US-0762377.
PR  22-JUL-1996; 96US-0684750.
XX
XX  (COGN-) COGNETIX INC.
PA
XX  Layer RT, McCabe RT, Zhou L;
PI
XX  WPI: 1998-120469/11.
DR  P-PSDB: AAW49989.
XX
XX  Use of conantokin peptide(s) - for treating disorders involving
PT  excessive excitation of nerve cells by excitatory amino acids or
PT  agonists of the N-methyl-D-aspartate receptor
XX
XX  Example 4; Pages 79-80; 122pp; English.
PS
XX
XX  The present sequence encodes Conus geographus conantokin, peptide
CC  derivatives of which can be used to treat neurologic and
CC  psychiatric disorders, e.g. as an anticonvulsant, neuroprotective
CC  or analgesic agent. Neurologic and psychiatric disorders include
CC  epilepsy, convulsions, neurotoxic injury (associated with
CC  conditions of hypoxia, anoxia or ischaemia, which typically follow
CC  stroke, cerebrovascular accident, brain or spinal cord trauma,
CC  myocardial infarct, physical trauma, drowning, suffocation,
CC  perinatal asphyxia or hypoglycaemic events), neurodegeneration
CC  associated with Alzheimer's disease, senile dementia, Amyotrophic
CC  lateral Sclerosis, Multiple Sclerosis, Parkinson's disease,
CC  Huntington's disease, Down's Syndrome, Korsakoff's disease,
CC  schizophrenia, AIDS dementia, multi-infarct dementia, Blauwanger
CC  dementia and neuronal damage associated with uncontrolled
CC  seizures), chemical toxicity (such as addiction, and morphine,
CC  opiates, opioid and barbiturate tolerance), pain (acute, chronic,
CC  migraine), anxiety, major depression, manic-depressive illness,
CC  obsessive-compulsive disorder, schizophrenia and mood disorders
CC  (such as bipolar disorder, unipolar depression, dysthymia and
CC  seasonal affective disorder) and dystonia (movement disorder),
CC  sleep disorder, muscle relaxation and urinary incontinence. The
CC  peptide can also be used to treat HIV infection, ophthalmic
CC  indication and memory, learning or cognitive defects.
XX
XX  Sequence 719 BP; 217 A; 171 C; 149 G; 182 T; 0 other;
SQ

```

Query Match 56.0%; Score 14; DB 19; Length 719;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 caggtgcatgcacca 15
|||||
DB 118 CAGGTGCTTCGCCA 105

```

AAC98831
ID  AAC98831 standard; CDNA; 1190 BP.
XX
XX  AAC98831;
AC
XX  09-MAR-2001 (first entry)
DF
XX
XX  Human pancreatic cancer antigen nucleotide sequence SEQ ID NO:59.
DE
XX
XX  Human; pancreas; pancreatic cancer; pancreatic cancer antigen;
KW  detection; diagnosis; identification; cytostatic; neuroprotective;
KW  neoplastic; immunomodulatory; relaxant; contraceptive; gynaecological;
KW  antiinflammatory; cardiant; gene therapy; chromosome mapping;
KW  linkage analysis; tissue identification; tissue typing; forensic;
KW  neural; immune system; muscular; reproductive; gastrointestinal;
KW  pulmonary; cardiovascular; renal; proliferative; ss.
XX
XX  Homo sapiens.
OS
XX
XX  WO200055320-A1.
PN
XX
XX  21-SEP-2000.
PD
XX
XX  08-MAR-2000; 2000WO-US05989.
PF
XX  12-MAR-1999; 99US-0124270.
PR
XX  (HUMA-) HUMAN GENOME SCI INC.
PA
XX
XX  Rosen CA, Ruben SM;
PI
XX  WPI: 2000-579444/54.
DR  P-PSDB: AAB54066.
XX
XX  New nucleic acid that is a pancreatic cancer antigen for preventing,
PT  treating, or ameliorating a medical condition, particular pancreatic
PT  cancer, or for use in assays for diagnosing a pathological condition -
XX
XX  Claim 1; Page 537; 1379pp; English.
PS
XX
XX  AAC98773 to AAC99231 encode the human pancreatic cancer associated
CC  proteins, called pancreatic cancer antigens, given in AAB54008 to
CC  AAB54466. The human pancreatic cancer antigens have cytostatic,
CC  neuroprotective, neoplastic, immunomodulatory, relaxant, contraceptive,
CC  gynaecological, cardiant and antiinflammatory activities, and can be used
CC  in gene therapy. The polynucleotide and proteins can be used for
CC  preventing, treating, or ameliorating a medical condition or in assays
CC  for diagnosing a pathological condition or a susceptibility to one in a
CC  subject. Binding partners to the proteins and the activity of the
CC  proteins can be identified. The pancreatic cancer antigens can be used to
CC  detect, treat or prevent pancreatic disorders, especially cancer.
CC  Agonists and antagonists to the antigens can be screened for. The
CC  pancreatic cancer antigen polynucleotides can be used to design nucleic
CC  acid hybridisation probes that can be used in chromosome mapping, linkage
CC  analysis, tissue identification and/or typing and a variety of forensic
CC  and diagnostic methods. The proteins can be used to generate antibodies
CC  which are used to purify, detect and target the polypeptides, including
CC  both in vivo and in vitro diagnostic and therapeutic methods. The
CC  proteins can be used to treat or prevent neural, immune system, muscular,
CC  reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
CC  proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
CC  sequences used in the exemplification of the present invention.
XX
XX  Sequence 1190 BP; 269 A; 323 C; 329 G; 267 T; 2 other;
SQ

```

Query Match 56.0%; Score 14; DB 21; Length 1190;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 ggtgcatgcaccac 17
|||||
DB 1141 ggtgcatgcaccac 1154

RESULT 13

AAZ56381/c
ID AAZ56381 standard; DNA: 1344 BP.

AC AAZ56381;

DT 17-MAR-2000 (first entry)

DE Escherichia coli flagellin protein nucleotide sequence SEQ ID NO:51.

XX flagellin; flhC; antigen; detection; ds.

OS Escherichia coli.

PN WO9961458-A1.

PD 02-DEC-1999.

PF 21-MAY-1999; 99WO-AU00385.

PR 21-MAY-1998; 98AU-0003634.

PA (UNSY) UNIV SYDNEY.

PI Reeves PR, Wang L;

DR WPI; 2000-072598/06.

PT Novel nucleic acid molecule useful for the detection of flagellated

XX bacterial strains in food, faeces, etc.

PS Claim 3; Page 225; 245pp; English.

AAZ56331 to AAZ56398 represent nucleic acid molecules (1) encoding all or part of an Escherichia coli flagellin protein except a protein expressed by E. coli H1, H7, H12 or H48 type strains. The present invention also describes a method of detecting the presence of E. coli of a particular H serotype in a sample, comprising specifically hybridizing a nucleic acid, preferably at least a pair, derived from a flagellating gene, specific for a particular flagellin gene associated with the H serotype, to any E. coli in the sample which contain the gene, and detecting any hybridised molecules, identifying the presence of that of E. coli of H serotype (1) are useful for: (1) detecting the presence of a pair of (1) to any E. coli in the sample and detecting the presence of both O and H-serotypes of E. coli by hybridising at least one or a pair of (1) to any E. coli present in the sample and detecting the hybridised combination of O and H antigen. Hybridised (1) when using at least one (1) is detected by southern blot analysis and, when using a pair of (1), is detected by polymerase chain reaction (PCR). AAZ56339 to AAZ56420 represent primers used in the exemplification of the present invention.

Sequence 1344 BP; 362 A; 337 C; 322 G; 323 T; 0 other;

Query Match

Best Local Similarity 56.0%; Score 14; DB 21; Length 1344;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 gcatcgccagcatt 20

Db 590 GCATCGCCAGCATT 577

RESULT 14

AAA69762/c

ID AAA69762 standard; cDNA; 2017 BP.

AC AAA69762;

XX 07-NOV-2000 (first entry)

XX Human ovarian carcinoma antigen polynucleotide SEQ ID NO:72.

DE Human ovarian carcinoma; ovarian cancer; therapy; diagnosis;

XX tumour antigen; identification; cytostatic; gene therapy; vaccine; ss.

OS Homo sapiens.

PN WO200036107-A2.

PD 22-JUN-2000.

PF 17-DEC-1999; 99WO-US30270.

PR 17-DEC-1998; 98US-0215681.

PR 17-DEC-1998; 98US-0216003.

PR 23-JUN-1999; 99US-0338933.

PR 24-SEP-1999; 99US-0404879.

PA (CORI-) CORIXA CORP.

PI Mitcham JL, King GE, Algate PA, Frudakis TN;

DR WPI; 2000-431589/37.

PT Immunogenic portion of an ovarian carcinoma protein and the nucleic

XX acid encoding it, useful for the diagnosis, prevention and treatment of

XX cancer, preferably ovarian cancer

PS Claim 1; Fig 2; 299pp; English.

The present invention describes an isolated polypeptide comprising an immunogenic portion of an ovarian carcinoma protein (or its variants). Ovarian carcinoma proteins, and polynucleotides encoding them, have cytostatic activity and can be used in gene therapy and vaccines. Ovarian carcinoma polypeptides, nucleic acids, antibodies and vaccines are useful for the prevention, diagnosis and treatment of cancer, preferably ovarian cancer. AAA69691 to AAA70077 and AAB12552 to AAB12557 represent human ovarian carcinoma polynucleotides and proteins used in the exemplification of the present invention.

Sequence 2017 BP; 673 A; 440 C; 462 G; 442 T; 0 other;

Query Match

Best Local Similarity 56.0%; Score 14; DB 21; Length 2017;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 agtgcacgcgcag 16

Db 964 AGTGCAATCGCCAG 951

RESULT 15

AAT30684/c

ID AAT30684 standard; DNA: 2193 BP.

AC AAT30684;

DT 11-SEP-1996 (first entry)

DE Kaposi's sarcoma associated herpesvirus ORF22.

XX Kaposi's sarcoma associated herpesvirus; KSHV; gamma-2 herpesvirus;

XX therapy; diagnosis; antisense; triplex; vaccine; AIDS; ss.

OS Kaposi's sarcoma associated herpesvirus.

PN WO9615779-A1.

PR 30-MAY-1996.

XX 21-NOV-1995; 95WO-US15138.
 XX 11-APR-1995; 95US-0420235.
 PR 21-NOV-1994; 94US-0343101.
 XX (UYCO) UNIV COLUMBIA NEW YORK.
 PA Chang Y, Moore PS;
 PI WPI: 1996-268320/27.
 DR P-PSDB; AAR97832.
 XX Herpes virus associated with Kaposi's sarcoma - also definitive DNA
 PT sequences, useful for diagnosis of and to develop prods. for
 treatment of Kaposi's sarcoma
 XX Claim 17; Page 198-201; 277pp; English.
 PS
 CC Lambda clone KS5 (AAT30681) has 17 protein-coding regions (AAT30682-98),
 CC 15 of which are complete ORFs, including ORF22 (AAT30684) (nts 2093-
 CC 4285), which codes for glycoprotein-H (AAR97832). KS5 is a fragment
 CC of a newly identified human gamma-2 herpesvirus associated with
 CC Kaposi's sarcoma (KS). The ORFs were identified by sequence and
 CC positional homology to known herpesvirus sequences. Probes that bind
 CC specifically to the KS associated herpesvirus can be used for KS
 CC diagnosis. Antisense or triplex oligonucleotides are useful for
 CC prophylaxis or treatment of KS, and the protein products (see also
 CC AAR97830-46) of the 17 open reading frames are useful as vaccines.
 XX
 SQ Sequence 2193 BP; 510 A; 597 C; 536 G; 550 T; 0 other;

Query Match 56.0%; Score 14; DB 17; Length 2193;
 Best Local Similarity 100.0%; Pred. No. 34;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 tgcacgcacgat 19
 |||||||
 DB 46 TGCATCGCCAGCAT 33

RESULT 16
 AAT16814/C
 ID AAT16814 standard; DNA: 2193 BP.
 XX
 AC AAT16814;
 XX
 DT 13-AUG-1996 (first entry)
 XX
 DE Kaposi's sarcoma associated herpesvirus ORF22.
 XX
 KW Kaposi's sarcoma; KSHV; lymphoma; AIDS; vaccine; diagnosis; therapy;
 KM glycoprotein H; ss.
 XX
 OS Kaposi's sarcoma associated herpesvirus.
 XX
 PN W09606159-A1.
 PD 29-FEB-1996.
 PF 11-AUG-1995; 95WO-US10194.
 XX
 PR 11-APR-1995; 95US-0420235.
 PR 18-AUG-1994; 94US-0292365.
 PR 21-NOV-1994; 94US-0343100.
 XX
 PA (UYCO) UNIV COLUMBIA NEW YORK.
 PA (GRANT) GRANT D E.
 PA (VIELE) VIELE L.
 XX
 PI Chang Y, Moore PS, Grant DE, Viele L;
 XX

DR WPI: 1996-151362/15.
 DR P-PSDB; AAR93608.
 XX
 PT Herpes virus DNA associated with Kaposi's sarcoma - also associated
 PT vectors and proteins, used in detection and vaccination.
 XX
 PS Claim 17; Page 211-214; 305pp; English.
 XX
 CC Kaposi's sarcoma associated herpes virus (KSHV) clone KS5 (AAT16806),
 CC obd. from a KS lesion genomic library, includes 15 complete ORFs and
 CC 2 incomplete ORFs (AAT16807-23) named according to their herpesvirus
 CC salmuri positional homologues. ORF22 (AAT16814) codes for
 CC glycoprotein H (AAR93608). KSHV DNA may be incorporated into a vector
 CC and expressed in host cells to produce peptides useful in vaccines or
 CC for raising antibodies. The DNA may itself be used to immunise a
 CC subject, or used to design therapeutic antisense and triplex
 CC molecules or diagnostic probes, or to raise transgenic animals.
 XX
 SQ Sequence 2193 BP; 510 A; 597 C; 536 G; 550 T; 0 other;

Query Match 56.0%; Score 14; DB 17; Length 2193;
 Best Local Similarity 100.0%; Pred. No. 34;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 tgcacgcacgat 19
 |||||||
 DB 46 TGCATCGCCAGCAT 33

RESULT 17
 AAT20914/C
 ID AAT20914 standard; DNA: 2193 BP.
 XX
 AC AAT20914;
 XX
 DT 02-DEC-1999 (first entry)
 XX
 DE Nucleotide sequence of gH glycoprotein of HHV-8.
 XX
 KW vaccine; antibody; Kaposi's sarcoma; glycoprotein; detection;
 KM HHV-8; Human Herpesvirus type 8; ds.
 XX
 OS Human Herpes Virus type 8.
 XX
 PN W09945948-A1.
 PD 16-SEP-1999.
 PF 12-MAR-1999; 99WO-US05464.
 XX
 PR 12-MAR-1998; 98US-0077670.
 XX
 PA (UYPE-) UNIV PENNSYLVANIA.
 PA Whitbeck JC, Eisenberg RJ, Cohen GH;
 PI WPI: 1999-551211/46.
 DR P-PSDB; AAY42395.
 XX
 PT New nucleic acid molecules encoding human herpesvirus type glycoprotein
 PT H and glycoprotein L polypeptides for diagnosis and treatment of
 PT Kaposi's Sarcoma -
 XX
 PS Claim 3; Fig 1; 75pp; English.
 XX
 CC This is the nucleotide sequence of the human herpes virus type 8
 CC glycoprotein H. This sequence and it's encoded protein (AAY42395) can be
 CC used in the development of vaccines, antibodies and other preparations.
 CC The vaccine is useful as a therapeutic agent for prevention of HHV-8
 CC disease and as a therapeutic agent for treatment of HHV-8 infection
 CC in order to boost the immune response in the animal. Administration of
 CC a pure preparation of soluble HHV-8 gH/gL complex or at least one

CC polynucleotide encoding the soluble complex is useful for preventing or
 CC treating an HIV-8 infection in a human.
 CC Diagnosing of an HIV-8 infection in a human is carried out by contacting
 CC a biological sample with an antibody which specifically binds to either
 CC glycoprotein or in complex form. If binding of the antibody occurs then
 CC the human has HIV-8 infection.
 CC Alternatively the infection is diagnosed by addition of one of the
 CC HIV-8 glycoproteins or complex and determining whether specific binding
 CC of the protein to an antibody in the biological sample occurs. A further
 CC diagnostic method involves contacting the biological sample with an
 CC HIV-8 glycoprotein H polynucleotide or HIV-8 glycoprotein L
 CC polynucleotide and determining whether binding to a nucleic acid in the
 CC sample occurs.

SO Sequence 2193 BP; 510 A; 597 C; 536 G; 550 T; 0 other;

Query Match 56.0%; Score 14; DB 20; Length 2193;
 Best Local Similarity 100.0%; Pred. No. 34;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 tgcacgcacgcat 19
 |||||
 Db 46 TGCATGCCGACGACAT 33

RESULT 18

AAZ39011/C
 ID AAZ39011 standard; cDNA; 3593 BP.

AC AAZ39011;

DT 28-FEB-2000 (first entry)

DE Mouse Ese2 coding sequence.

KW Mouse; murine; Ese1; Ese2; endocytosis; vesicular trafficking;
 regulation; actin cytoskeleton; detection; cancer; infection;
 EH-domain and SH3-domain regulator of endocytosis; anticancer;
 antiproliferative; antiviral; ss.

OS Mus sp.

PN WO955728-A2.

PD 04-NOV-1999.

PF 27-APR-1999; 99WO-CA00375.

PR 27-APR-1998; 98CA-2230201.

PR 05-FEB-1999; 99US-0118739.

PA (HSCR-) HSC RES & DEV LP.

PI Egan SE, Wang W, Sengar A.

DR WPI: 2000-052802/04.

DR P-PSDB: AAY57445.

PT New nucleic acid encoding Ese1 and 2 proteins, involved in regulation
 of endocytosis, used e.g. for treating cancer or preventing viral
 infection

PT Claim 25; Page 46-48; 99pp; English.

PS The present sequence encodes mouse Ese2. The present invention
 CC specifically describes mammalian Ese1 and 2 proteins (I) and their splice
 CC variants (Ese = EH-domain and SH3-domain regulator of endocytosis). (I)
 CC are involved in regulation of clathrin-mediated endocytosis (as a complex
 CC with Eps15 protein), vesicular trafficking and actin cytoskeleton.
 CC Generally (I) (or its (ant)agonists, mimetics, fragments and inactive
 CC mutants); (I)-specific antibodies (Ab); sequences antisense to the (I)
 CC polynucleotide; agents that downregulate expression of Ese genes or

CC antagonists of an Ese binding partner are used to treat diseases
 CC associated with undesirable endocytosis and resulting changes in cellular
 CC function. Particularly overexpression of Ese1 is used to block
 CC clathrin-mediated endocytosis in vivo or in cell cultures, while
 CC administration of (I) is used to promote endocytosis of selected cells.
 CC (Ant)agonists of (I) or Ab are used to suppress abnormal proliferation of
 CC cells that can be stimulated to proliferate by a growth factor receptor;
 CC and similar compounds (also inactive Ese mutants) can be used to prevent
 CC viral infection. Endocytosis may also be regulated, in vivo or in cell
 CC cultures, by forming an Ese-Eps15 complex, then binding dynamin to the
 CC complex. Generally conditions that can be treated include cancer;
 CC abnormal cell division or migration; viral infection; or abnormal
 CC receptor signalling, tissue development or synaptic transmission.

SO Sequence 3593 BP; 1171 A; 783 C; 850 G; 789 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 3593;
 Best Local Similarity 100.0%; Pred. No. 33;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 aggtgcatgcacg 16
 |||||
 Db 923 AGGTGCATGCCACG 910

RESULT 19

AAZ49334
 ID AAZ49334 standard; cDNA; 4189 BP.

AC AAZ49334;

DT 14-MAR-2000 (first entry)

DE Murine multidrug resistance-1 (MDR-1) cDNA.

KW Multidrug resistance; MDR-1; P-glycoprotein;
 transmembrane efflux pump; haematopoietic stem cell; transduction;
 bone marrow transplantation; chemotherapy; radiation therapy; cancer;
 gene therapy; gene replacement; genetic defect; thalassemia;
 Gaucher's disease; sickle cell anaemia; leukaemia; ex vivo expansion;
 cytokine; ds.

OS Mus sp.

PN WO9961589-A2.

PD 02-DEC-1999.

PF 27-MAY-1999; 99WO-US11825.

PR 28-MAY-1998; 98US-0086988.

PA (STUD-) ST JUDE CHILDREN'S RES HOSPITAL.

PI Sorrentino B, Bunting K;

DR WPI: 2000-072615/06.

DR P-PSDB: AAY58188.

PT Ex vivo expansion of hematopoietic stem cells transduced with a
 sequence encoding human multidrug resistance-1, used for bone marrow
 transplantation

PT Disclosure; Page 90-93; 113pp; English.

PS This sequence represents cDNA encoding murine multidrug
 CC resistance protein MDR-1. MDR-1 is a transmembrane

CC efflux pump, responsible for the export of drugs from cells,
 CC particularly cancer cells. The invention relates to transducing
 CC hematopoietic stem cells with nucleic acid encoding an MDR protein
 CC and culturing the modified cells. The modified hematopoietic stem
 CC cells are useful in bone marrow transplantation (to reconstitute
 CC hematopoietic systems in patients who have undergone chemotherapy or
 CC radiation therapy) and in ex vivo gene therapy of genetic defects in
 CC cells derived from hematopoietic stem cells, e.g., thalassemia,
 CC Gaucher's disease, sickle cell anaemia or leukaemia. The modified
 CC cells can also be used to identify factors involved in regulating
 CC proliferation and differentiation in hematopoietic stem cells.
 CC Hematopoietic stem cells that express MDR-1 will be protected against
 CC chemotherapeutic agents, so can be engrafted while the patient is
 CC undergoing chemotherapy. Expansion of (rare) hematopoietic stem cells
 CC provides sufficient cells to permit standard biochemical analysis.
 CC Overexpression of MDR-1 allows cytokine-driven expansion of
 CC hematopoietic stem cells by at least 10-fold compared with a maximum
 CC of 4-fold in known procedures.

XX Sequence 4189 BP; 1204 A; 875 C; 1028 G; 1082 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 4189;

Best Local Similarity 100.0%; Pred. No. 33;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 gccagcattcgat 25

Db 880 gccagcattcgat 893

RESULT 20

AAQ38950

ID AAQ38950 standard; DNA; 4313 BP.

XX AAQ38950;

DT 28-JUL-1993 (first entry)

DE Mouse multidrug resistance sequence.

KW mdr gene; lambda DR11 clone; ss.

OS Mus musculus.

XX Key Location/Qualifiers

FT CDS 110..3940

FT /*tag= a

FT /phenotype= multidrug_resistance

XX US198344-A.

PD 30-MAR-1993.

PF 15-JUL-1986; 86US-0885951.

PR 15-JUL-1986; 86US-0885951.

PR 06-FEB-1991; 91US-0652311.

XX (MASI) MASSACHUSETTS INST TECHNOLOGY.

PI Croop JM, Gros P, Housman DE;

DR WPI; 1993-126077/15.

DR P-PSDB; AAR35199.

XX DNA sequence which confers multi-drug resistance on sensitive

XX mammalian cells - used to preserve bone marrow cells during

XX chemotherapy to prevent infection

XX Claim 1; Fig 8; 22pp; English.

XX A cDNA library was constructed from mRNA which had been isolated

CC from a drug-sensitive mouse cell-line. Two mouse cDNA molecules
 CC complementary to the mRNA species encoded by 2 related but distinct
 CC cDNA genes were isolated and cloned. One of the cDNA clones (lambda
 CC DR11) is a full-length cDNA clone for one member of the mdr gene
 CC family. When incorporated into prokaryotic expression vector pUREX4
 CC (which allows high levels of transcription of the cDNA when
 CC introduced into mammalian cells), the clone was shown to confer the
 CC multidrug resistance phenotype upon transfection into drug-sensitive
 CC mammalian cells.

XX Sequence 4313 BP; 1241 A; 904 C; 1058 G; 1110 T; 0 other;

Query Match 56.0%; Score 14; DB 14; Length 4313;

Best Local Similarity 100.0%; Pred. No. 33;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 gccagcattcgat 25

Db 989 gccagcattcgat 1002

RESULT 21

AAZ39010/C

ID AAZ39010 standard; cDNA; 4625 BP.

XX AAZ39010;

DT 28-FEB-2000 (first entry)

DE Mouse Ese2 full length cDNA sequence.

KW Mouse; murine; Esel; Ese2; endocytosis; vesicular trafficking;

KW regulation; actin cytoskeleton; detection; cancer; infection;

KW EH-domain and SH3-domain regulator of endocytosis; anticancer;

XX antiproliferative; antiviral; ss.

OS Mus sp.

PN W09955728-A2.

PD 04-NOV-1999.

PF 27-APR-1999; 99WO-CA00375.

PR 27-APR-1998; 98CA-2230201.

PR 05-FEB-1999; 99US-0118739.

XX (HSCR-) HSC RES & DEV LP.

PI Egan SE, Wang W, Sengar A;

DR WPI; 2000-052802/04.

DR P-PSDB; AAY57445.

XX New nucleic acid encoding Esel and 2 proteins, involved in regulation

XX of endocytosis, used e.g. for treating cancer or preventing viral

XX infection

XX Claim 25; Page 43-46; 99pp; English.

XX The present sequence encodes mouse Ese2. The present invention

XX specifically describes mammalian Esel and 2 proteins (I) and their splice

XX variants (Ese = EH-domain and SH3-domain regulator of endocytosis). (I)

XX are involved in regulation of clathrin-mediated endocytosis (as a complex

XX with Esp15 protein), vesicular trafficking and actin cytoskeleton.

XX Generally (I) (or its (ant)agonists, mimetics, fragments and inactive

XX mutants): (I)-specific antibodies (Ab); sequences antisense to the (I)

XX polynucleotide; agents that downregulate expression of Ese genes or

XX antagonists of an Ese binding partner are used to treat diseases

XX associated with undesirable endocytosis and resulting changes in cellular

XX function. Particularly overexpression of Esel is used to block

XX clathrin-mediated endocytosis in vivo or in cell cultures, while

CC administration of (I) is used to promote endocytosis of selected cells.
 CC (Ant)agonists of (I) or Ab are used to suppress abnormal proliferation of
 CC cells that can be stimulated to proliferate by a growth factor receptor;
 CC and similar compounds (also inactive Ese mutants) can be used to prevent
 CC viral infection. Endocytosis may also be regulated, in vivo or in cell
 CC cultures, by forming an Ese-Esp15 complex, then binding dynamin to the
 CC complex. Generally conditions that can be treated include cancer;
 CC abnormal cell division or migration; viral infection; or abnormal
 CC receptor signalling, tissue development or synaptic transmission.
 CC
 SQ Sequence 4625 BP; 1453 A; 1009 C; 1104 G; 1059 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 4625;
 Best Local Similarity 100.0%; Pred. No. 33;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 3 aggtgcacgcag 16
 DB 1173 AGGTGATCGCCAG 1160

RESULT 22
 AAZ39027/C
 ID AAZ39027 standard; CDNA; 4975 BP.
 XX
 AC AAZ39027;
 XX
 DT 28-FEB-2000 (first entry)
 XX
 DE Mouse Ese2L coding sequence.
 XX
 KW Mouse; murine; Ese1; Ese2; endocytosis; vesicular trafficking;
 KW regulation; actin cytoskeleton; detection; cancer; infection;
 KW EH-domain and SH3-domain regulator of endocytosis; anticancer;
 KW antiproliferative; antiviral; ss.
 XX
 OS Mus sp.
 XX
 PN WO995728-A2.
 XX
 PD 04-NOV-1999.
 XX
 PF 27-APR-1999; 99MO-CA00375.
 XX
 PR 27-APR-1998; 98CA-2230201.
 PR 05-FEB-1999; 99US-0118739.
 XX
 PA (HSCR-) HSC RES & DEV LP.
 XX
 PI Egan SE, Wang W, Sengar A;
 XX
 DR WPI: 2000-052802/04.
 DR P-PSDB; AAY57450.
 XX
 PT New nucleic acid encoding Ese1 and 2 proteins, involved in regulation
 PT of endocytosis, used e.g. for treating cancer or preventing viral
 PT infection -
 XX
 PS
 XX

Claim 25; Page 66-69; 99pp; English.

CC The present invention specifically describes mammalian Ese1 and 2
 CC proteins (I) and their splice variants (Ese - EH-domain and SH3-domain
 CC regulator of endocytosis). (I) are involved in regulation of clathrin-
 CC mediated endocytosis (as a complex with Esp15 protein), vesicular
 CC trafficking and actin cytoskeleton. Generally (I) (or its (ant)agonists,
 CC mimetics, fragments and inactive mutants); (I)-specific antipodols (Ab);
 CC sequences antisense to the (I) polynucleotide; agents that downregulate
 CC expression of Ese genes or antagonists of an Ese binding partner are
 CC used to treat diseases associated with undesirable endocytosis and
 CC resulting changes in cellular function. Particularly overexpression of
 CC Ese1 is used to block clathrin-mediated endocytosis in vivo or in cell
 CC cultures, while administration of (I) is used to promote endocytosis of

CC selected cells. (Ant)agonists of (I) or Ab are used to suppress abnormal
 CC proliferation of cells that can be stimulated to proliferate by a growth
 CC factor receptor; and similar compounds (also inactive Ese mutants) can be
 CC used to prevent viral infection. Endocytosis may also be regulated, in
 CC vivo or in cell cultures, by forming an Ese-Esp15 complex, then binding
 CC dynamin to the complex. Generally conditions that can be treated include
 CC cancer; abnormal cell division or migration; viral infection; or abnormal
 CC receptor signalling, tissue development or synaptic transmission. The
 CC present sequence represents mouse Ese2L coding sequence.
 CC
 SQ Sequence 4975 BP; 1540 A; 1166 C; 1201 G; 1068 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 4975;
 Best Local Similarity 100.0%; Pred. No. 33;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 3 aggtgcacgcag 16
 DB 923 AGGTGATCGCCAG 910

RESULT 23
 AAV35235/C
 ID AAV35235 standard; DNA; 5319 BP.
 XX
 AC AAV35235;
 XX
 DT 28-SEP-1998 (first entry)
 XX
 DE R. prowazekii S-layer protein genomic DNA.
 XX
 KW Surface layer protein; S-layer; vaccine; typhus; spotted fever;
 KW infection; diagnosis; disease; ds.
 XX
 OS Rickettsia prowazekii.
 XX
 FH Key Location/Qualifiers
 FH -35_signal 340..345
 FT /*tag= a
 FT 363..368
 FT /*tag= b
 FT 379..386
 FT /*tag= c
 FT 391..5229
 FT /*tag= d
 FT /product= S-layer protein
 FT 5270..5306
 FT /*tag= e
 XX
 PN US5783441-A.
 XX
 PD 21-JUL-1998.
 XX
 PF 20-DEC-1993; 93US-0169927.
 XX
 PR 20-DEC-1993; 93US-0169927.
 PR 09-AUG-1991; 91US-0742128.
 XX
 PA (USNA) US SEC OF NAVY.
 XX
 PI Carl M, Ching W, Dasch GA, Dobson ME;
 XX
 DR WPI: 1998-427031/36.
 DR P-PSDB; AAW65088.
 XX
 PT Recombinant DNA encoding Rickettsia surface layer proteins - useful
 PT for diagnosing typhus and spotted fever and for preparing vaccines
 PT against them
 XX
 PS Claim 1; Column 11-24; 20pp; English.
 XX
 CC This sequence encodes the Surface layer (S-layer) protein from

CC R. prowazekii strain Breinl. This sequence is useful for vaccination
CC against typhus and spotted fever rickettsial infection or for diagnosing
CC diseases caused by these bacteria. The surface layer protein antigens can
CC be produced recombinantly in large quantities.

XX Sequence 5319 BP; 1815 A; 766 C; 983 G; 1755 T; 0 other;

Query Match 56.0%; Score 14; DB 19; Length 5319;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 8 catgccagcattt 21
|||||
Db 2221 CATGCCAGCATTT 2208

RESULT 24
AAV21186/c
ID AAV21186 standard; DNA: 5676 BP.

XX AAV21186;
XX
XX 24-JUL-1998 (first entry)

XX Amycolatopsis mediterranei strain Wt3136 5.7 kb KpnI DNA fragment.

XX Amycolatopsis mediterranei: rifamycin; synthesis: gene cluster;
XX polypeptide synthase; actinomycete: ansamycin; ds.

XX Amycolatopsis mediterranei.

XX Key Location/Qualifiers
XX CDS 3..5676
XX /*tag= a
XX /product= "polyketide synthase"
XX /note= "no stop codon given"

XX WO9807868-A1.

XX 26-FEB-1998.

XX 18-AUG-1997; 97WO-EP04495.

XX 20-AUG-1996; 96EP-0810551.

XX (NOVS) NOVARTIS AG.

XX Engel N, Schnupp T, Toupet C;

XX WPI: 1998-169172/15.

XX P-PSDB: AAW52844.

XX Amycolatopsis mediterranei rifamycin synthesis gene cluster - used
XX to produce rifamycin and rifamycin analogues

XX Claim 4; Page 38-43; 205pp; English.

XX The present sequence represents Amycolatopsis mediterranei strain Wt3136
XX 5.7 kb KpnI fragment DNA, from the present invention. The present
XX invention describes a Amycolatopsis mediterranei rifamycin synthesis
XX gene cluster DNA fragment comprising a DNA region involved directly or
XX indirectly in the gene cluster responsible for rifamycin synthesis,
XX including the adjacent DNA regions to the right and left which, by
XX reason of their function in connection with rifamycin biosynthesis,
XX qualify as constituents of this rifamycin gene cluster, and functional
XX fragments, derivatives or constituents of these. The Amycolatopsis
XX mediterranei rifamycin synthesis gene cluster DNA fragment can be used
XX for producing rifamycin, rifamycin analogues or precursors. It can also
XX be used for inactivating or modifying genes involved in ansamycin or
XX rifamycin biosynthesis. The DNA can be used for constructing mutant
XX actinomycetes strains from which the natural rifamycin or ansamycin
XX biosynthesis gene cluster has been partly or completely deleted. The

CC DNA fragment can be used for assembling a library of polyketide
CC synthases, which can be used for assembling a library of polyketides.
CC A hybridisation probe of the invention can be used for identifying DNA
CC fragments involved in the biosynthesis of ansamycins.

XX Sequence 5676 BP; 674 A; 1990 C; 2205 G; 807 T; 0 other;

Query Match 56.0%; Score 14; DB 19; Length 5676;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 caggtgcacgcaca 15
|||||
Db 3923 CAGGTGCATCGCCA 3910

RESULT 25
AAZ39026/c
ID AAZ39026 standard; cDNA: 6014 BP.

XX AAZ39026;

XX 28-FEB-2000 (first entry)

XX Mouse Ese2L cDNA sequence.

XX Mouse; murine; Ese2; endocytosis; vesicular trafficking;
XX regulation; actin cytoskeleton; detection; cancer; infection;
XX EH-domain and SH3-domain regulator of endocytosis; anticancer;
XX antiproliferative; antiviral; ss.

XX Mus sp.

XX WO9955728-A2.

XX 04-NOV-1999.

XX 27-APR-1999; 99WO-CA00375.

XX 27-APR-1998; 98CA-2230201.

XX 05-FEB-1999; 99US-0118739.

XX (HSCR-) HSC RES & DEV LP.

XX Egan SE, Wang W, Sengar A;

XX WPI: 2000-052802/04.

XX P-PSDB: AAY57450.

XX New nucleic acid encoding Ese1 and 2 proteins, involved in regulation
XX of endocytosis, used e.g. for treating cancer or preventing viral
XX infection

XX Claim 25; Page 63-66; 99pp; English.

XX The present invention specifically describes mammalian Ese1 and 2
XX proteins (I) and their splice variants (Ese - EH-domain and SH3-domain
XX regulator of endocytosis). (I) are involved in regulation of clathrin-
XX mediated endocytosis (as a complex with Eps15 protein), vesicular
XX trafficking and actin cytoskeleton. Generally (I) (or its (ant)agonists,
XX mimetics, fragments and inactive mutants); (I)-specific antibodies (Ab);
XX sequences antisense to the (I) polynucleotide; agents that downregulate
XX expression of Ese genes or antagonists of an Ese binding partner are
XX used to treat diseases associated with undesirable endocytosis and
XX resulting changes in cellular function. Particularly overexpression of
XX Ese1 is used to block clathrin-mediated endocytosis in vivo or in cell
XX cultures, while administration of (I) is used to promote endocytosis of
XX selected cells. (Ant)agonists of (I) or Ab are used to suppress abnormal
XX proliferation of cells that can be stimulated to proliferate by a growth
XX factor receptor, and similar compounds (also inactive Ese mutants) can be
XX used to prevent vital infection. Endocytosis may also be regulated, in
XX vivo or in cell cultures, by forming an Ese-Eps15 complex, then binding

CC dynamin to the complex. Generally conditions that can be treated include
 CC cancer; abnormal cell division or migration; viral infection; or abnormal
 CC receptor signaling, tissue development or synaptic transmission. The
 CC present sequence represents mouse Ese2L cDNA sequence.
 XX
 SQ Sequence 6014 BP: 1833 A; 1389 C; 1454 G; 1338 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 6014;
 Best Local Similarity 100.0%; Pred. No. 33;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 3 aggtcgcgcagcag 16
 |||||
 Db 1173 AGGTGATCGCCAG 1160

RESULT 26
 AAT30681/C
 ID AAT30681 standard; DNA: 20710 BP.

AC AAT30681;

DT 11-SEP-1996 (first entry)

DE Kaposi's sarcoma associated herpesvirus clone KS5.

KM Kaposi's sarcoma associated herpesvirus; KSHV; gamma-2 herpesvirus;

XX therapy; diagnosis; antisense; triplex; vaccine; AIDS; ss.

OS Kaposi's sarcoma associated herpesvirus.

XX
 FH Key Location/Qualifiers
 FT CDS complement (1..364)

FT /*tag= a
 FT /note= "incomplete ORF20"

FT 363..2105
 FT /*tag= b
 FT /product= thymidine kinase

FT /note= "ORF21"

FT 2093..4285

FT /*tag= c
 FT /product= glycoprotein H

FT /note= "ORF22"

FT complement (4282..5496)

FT /*tag= d
 FT /note= "ORF23"

FT complement (5500..7758)

FT /*tag= e
 FT /note= "ORF24"

FT 7757..11887

FT /*tag= f
 FT /note= "ORF25"

FT 11913..12830

FT /*tag= g
 FT /product= virion polypeptide VP23

FT /note= "ORF26"

FT 12851..13723

FT /*tag= h
 FT /product= minor capsid protein

FT /note= "ORF27"

FT 13969..14331

FT /*tag= i
 FT /note= "ORF28"

FT complement (14343..15485)

FT /*tag= j
 FT /product= (with ORF29A) putative DNA packaging

FT /note= "ORF29B"

FT 15609..15842

FT /*tag= k
 FT /note= "ORF30"

FT 15839..16423

FT CDS

FT /*tag= l
 FT /note= "ORF31"

FT 16390..17754

FT /*tag= m
 FT /note= "ORF32"

FT 17747..18685

FT /*tag= n
 FT /note= "ORF33"

FT complement (18743..19662)

FT /*tag= o
 FT /product= (with ORF29B) putative DNA packaging

FT /note= "ORF29A"

FT 19661..20644

FT /*tag= p
 FT /note= "ORF34"

FT 20625..20710

FT /*tag= q
 FT /note= "incomplete ORF35"

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

FT CDS

PS Claim 16; Page 164-175; 277pp; English.
 XX Phage lambda clone KS5 (AAT30681) was isolated from an AIDS-Kaposi's
 CC sarcoma (KS) genomic DNA library by screening with a specific DNA,
 CC KS30Bam, found in KS lesions. Sequence analysis of KS5 showed
 CC that KS30Bam is part of a larger herpesvirus genome. This agent
 CC was defined as a human gamma-2 herpesvirus that can be continuously
 CC cultured in naturally-transformed, EBV-coinfected lymphocytes from
 CC AIDS-related body-cavity based lymphomas. KS5 has 17 protein-coding
 CC regions (see also AAT30682-98) identified by homology to known gamma-
 CC herpesvirus sequences. Probes that bind specifically to the KS
 CC associated herpesvirus are useful for KS diagnosis. Antisense or
 CC triplex oligonucleotides are useful for prophylaxis or treatment of
 CC KS, and the peptide products (see also AAR37830-46) of the open reading
 CC frames are useful as vaccines.
 XX
 SQ Sequence 20710 BP: 4628 A; 5526 C; 5666 G; 4890 T; 0 other;

Query Match 56.0%; Score 14; DB 17; Length 20710;
 Best Local Similarity 100.0%; Pred. No. 32;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 6 tgcacgcgcagcat 19
 |||||
 Db 2138 TGCATCGCCAGCAT 2125

RESULT 27
 AAT16806/C
 ID AAT16806 standard; DNA: 20710 BP.

AC AAT16806;

DT 13-AUG-1996 (first entry)
 XX Kaposi's sarcoma associated herpesvirus clone KS5.
 DE Kaposi's sarcoma; KSHV; Lymphoma; AIDS; vaccine; diagnosis; therapy;
 XX KS.
 KM Kaposi's sarcoma associated herpesvirus.
 OS WO9606159-A1.
 PN 29-FEB-1996.
 PD 11-AUG-1995; 95WO-0510194.
 PF 11-APR-1995; 95US-0420235.
 PR 18-AUG-1994; 94US-0292365.
 PR 21-NOV-1994; 94US-0343100.
 XX (UYCO) UNIV COLUMBIA NEW YORK.
 PA (GRANT/) GRANT D E.
 PA (VIEL/) VIELE L.
 XX Chang Y, Moore PS, Grant DE, Viele L;
 PI WPI: 1996-151362/15.
 DR Herpesvirus DNA associated with Kaposi's sarcoma - also associated
 XX vectors and proteins, used in detection and vaccination.
 PT Claim 16: Page 177-188; 305pp; English.
 PS Representational difference analysis was used to identify unique
 XX DNA sequences in Kaposi's sarcoma (KS) tissue. An isolated sequence,
 CC KS3308am (AAT16824), was used to screen a KS lesion genomic library. A
 CC lambda phage, KS5, was identified and the sequence of its insert
 CC (AAT16808) was determined. This represents a novel infectious agent
 CC associated with AIDS-KS, Kaposi's sarcoma associated herpesvirus (KSHV).
 CC The KS5 fragment has 15 complete ORFs and 2 incomplete ORFs (see also
 CC AAT16807-23). Portions of the sequence may be incorporated into a
 CC vector and expressed in host cells to produce peptides (see also
 CC AAT93601-17) useful in vaccines or for raising antibodies. The DNA may
 CC itself be used to immunise a subject, or used to design therapeutic
 CC antisense and triplex molecules or diagnostic probes, or to raise
 CC transgenic animals.
 CC SQ Sequence 20710 BP; 4628 A; 5526 C; 5666 G; 4890 T; 0 other;

Query Match 56.0%; Score 14; DB 17; Length 20710;
 Best Local Similarity 100.0%; Pred. No. 32;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 tgcacgcacagcat 19
 ||||||||||||
 DB 2138 TGCATCGCCAGCAT 2125

RESULT 28
 AAV73803/C
 ID AAV73803 standard; DNA: 35100 BP.
 XX AAV73803;
 AC 25-FEB-1999 (first entry)
 XX KSHV LUR DNA (nucleotides 35,101-70,200).
 DE Kaposi's sarcoma; acquired immune deficiency syndrome; AIDS; DHFR; Bcl-2;
 KM dihydrofolate reductase; LUR; long unique region; vaccine; prophylaxis;
 KM diagnosis; treatment; HHV8; thymidine kinase; glycoprotein H;
 KM capsid protein; packaging protein; helicase primase; glycoprotein M;
 KM viral protein kinase; alkaline exonuclease; virion assembly protein;
 XX FH

KM uracil DNA glycosylase; UDG; glycoprotein L; ds.
 XX Kaposi's sarcoma-associated herpesvirus.
 OS US5849564-A.
 PN 15-DEC-1998.
 PD 29-NOV-1996; 96US-0770379.
 PF 29-NOV-1996; 96US-0770379.
 PR 29-NOV-1996; 96US-0770379.
 XX (UYCO) UNIV COLUMBIA NEW YORK.
 PA Bohenzky RA, Chang Y, Edelman IS, Moore PS, Russo JJ;
 PI WPI: 1999-069741/06.
 DR Kaposi's sarcoma-associated herpes virus nucleic acid - encodes
 PT dihydrofolate reductase and is useful for treatment, prophylaxis
 PT or diagnosis of Kaposi's sarcoma
 XX Disclosure; Column 97-126; 109pp; English.
 XX This sequence is a fragment of the Kaposi's sarcoma-associated
 CC herpesvirus (KSHV) LUR (long unique region). This fragment contains
 CC coding regions for ORF20, ORF21 which encodes thymidine kinase, ORF22
 CC which encodes glycoprotein H, ORF23, ORF24, ORF25 which encodes the
 CC major capsid protein, ORF26 which encodes capsid protein II, ORF27,
 CC ORF28, ORF29b which encodes packaging protein II, ORF30, ORF31, ORF32,
 CC ORF33, ORF29a which encodes packaging protein I, ORF34, ORF35, ORF36
 CC which encodes viral protein kinase, ORF37 which encodes alkaline
 CC exonuclease, ORF38, ORF39 which encodes glycoprotein M, ORF40 which
 CC encodes helicase primase subunit I, ORF41 which encodes helicase primase
 CC subunit II, ORF42, ORF43 which encodes capsid protein III, ORF44 which
 CC encodes helicase primase subunit III, ORF45 which encodes the virion
 CC assembly protein, ORF46 which encodes uracil DNA glycosylase (UDG),
 CC ORF47 which encodes glycoprotein L. KSHV is a new human Herpesvirus
 CC (HHV8) believed to cause Kaposi's sarcoma (KS) which is the most common
 CC form of neoplasm occurring in persons with acquired immune deficiency
 CC syndrome (AIDS). The DHFR protein is useful for vaccination, prophylaxis,
 CC diagnosis and treatment of a subject with Kaposi's sarcoma and for
 CC detecting expression of a DNA virus associated with Kaposi's sarcoma in a
 CC cell.
 CC SQ Sequence 35100 BP; 8016 A; 9231 C; 9477 G; 8376 T; 0 other;

Query Match 56.0%; Score 14; DB 20; Length 35100;
 Best Local Similarity 100.0%; Pred. No. 32;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 tgcacgcacagcat 19
 ||||||||||||
 DB 2058 TGCATCGCCAGCAT 2045

RESULT 29
 AAV21187/C
 ID AAV21187 standard; DNA: 53789 BP.
 XX AAV21187;
 AC 24-JUL-1998 (first entry)
 XX Amycolatopsis mediterranei rifamycin synthesis gene cluster fragment.
 DE Amycolatopsis mediterranei; rifamycin; synthesis; gene cluster;
 KM polyketide synthase; actinomycete; ansamycin; ds.
 XX Amycolatopsis mediterranei.
 OS Key Location/Qualifiers
 XX FH

PT the presence of a sequence polymorphism -
XX
PS Claim 1; Page 269; 543pp; English.
XX
CC Sequences AAA76318-A77509 represent 1192 human nucleic acid sequences
CC which contain single nucleotide polymorphisms (SNPs). Sequences 1 to
CC 1112 (AAA76318-A77429) are consecutive pairs of nucleotides which
CC contain silent SNPs. Sequences 1113 to 1192 (AAA77430-A77509) are
CC consecutive pairs of nucleotides containing SNPs which result in changes
CC in the corresponding amino acid sequences (AAB11749-B11828). The SNPs in
CC sequences 1113 to 1128 (AAA77430-A77445) lead to conservative amino acid
CC changes, while those in sequences 1129 to 1186 (AAA77446-A77503) result
CC in non-conservative changes. The SNPs in sequences 1187 to 1192
CC (AAA77504-A77509) generate frameshift mutations. The invention also
CC relates to a method of detecting a polymorphic site in a nucleic acid and
CC a method of determining the relatedness of two nucleic acids. It also
CC encompasses peptides containing polymorphic sites, antibodies raised
CC against such peptides, and a method of detecting polymorphic
CC proteins/peptides using the antibodies. The nucleic acids are useful for
CC gene therapy of an individual having, suspected of having, or at risk of
CC developing a pathological condition due to the presence of a sequence
CC polymorphism. Such treatment would comprise administration of the
CC wild-type nucleic acid sequence. Antibodies raised against polymorphic
CC peptides can also be used in the treatment of such individuals.
SQ
Sequence 51 BP; 13 A; 11 C; 15 G; 12 T; 0 other;
Query Match 52.0%; Score 13; DB 21; Length 51;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 11 cgcgcagcatctcg 23
DB 11 cgcgcagcatctcg 23
RESULT 33
AAA76687
ID AAA76687 standard; cDNA; 51 BP.
XX
AC AAA76687;
XX
DT 16-NOV-2000 (first entry)
XX
DE Human clone cg28389525 polymorphic site, SEQ ID NO:370.
XX
KW Human; single nucleotide polymorphism; SNP;
KW detection; identification; gene therapy; ss.
XX
OS Homo sapiens.
FH Key Location/Qualifiers
FT Variation replace (26,T)
FT /*tag= a
XX
PN WO200029623-A2.
XX
PD 25-MAY-2000.
XX
PF 17-NOV-1999; 99WO-US27293.
XX
PR 17-NOV-1998; 98US-0109024.
PR 16-NOV-1999; 99US-0109024.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shinkels RA, Leach MD;
XX
DR MPI: 2000-387826/33.
XX
PT Human nucleic acids containing single nucleotide polymorphisms, useful
PT for treating a subject suffering, or at risk from a pathology due to

PT the presence of a sequence polymorphism -
XX
PS Claim 1; Page 270; 543pp; English.
XX
CC Sequences AAA76318-A77509 represent 1192 human nucleic acid sequences
CC which contain single nucleotide polymorphisms (SNPs). Sequences 1 to
CC 1112 (AAA76318-A77429) are consecutive pairs of nucleotides which
CC contain silent SNPs. Sequences 1113 to 1192 (AAA77430-A77509) are
CC consecutive pairs of nucleotides containing SNPs which result in changes
CC in the corresponding amino acid sequences (AAB11749-B11828). The SNPs in
CC sequences 1113 to 1128 (AAA77430-A77445) lead to conservative amino acid
CC changes, while those in sequences 1129 to 1186 (AAA77446-A77503) result
CC in non-conservative changes. The SNPs in sequences 1187 to 1192
CC (AAA77504-A77509) generate frameshift mutations. The invention also
CC relates to a method of detecting a polymorphic site in a nucleic acid and
CC a method of determining the relatedness of two nucleic acids. It also
CC encompasses peptides containing polymorphic sites, antibodies raised
CC against such peptides, and a method of detecting polymorphic
CC proteins/peptides using the antibodies. The nucleic acids are useful for
CC gene therapy of an individual having, suspected of having, or at risk of
CC developing a pathological condition due to the presence of a sequence
CC polymorphism. Such treatment would comprise administration of the
CC wild-type nucleic acid sequence. Antibodies raised against polymorphic
CC peptides can also be used in the treatment of such individuals.
SQ
Sequence 51 BP; 13 A; 12 C; 15 G; 11 T; 0 other;
Query Match 52.0%; Score 13; DB 21; Length 51;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 11 cgcgcagcatctcg 23
DB 11 cgcgcagcatctcg 23
RESULT 34
AAA45502/c
ID AAA45502 standard; cDNA; 286 BP.
XX
AC AAA45502;
XX
DT 21-AUG-2000 (first entry)
XX
DE Human secreted expressed sequence tag SEQ ID NO:2077.
XX
KW Human; mouse; chicken; rat; secreted expressed sequence tag; SEST;
KW expressed sequence tag; EST; probe; chemotactic; proliferative;
KW immunomodulatory; haematopoietic; chemokine; analgesic; haemostatic;
KW thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal;
KW antiviral; antidiabetic; antiasthmatic; vulnereary; antiparkinsonian;
KW cerebroprotective; neuroprotective; antidiabetic; antiparkinsonian;
KW vaccine; autoimmune disorder; anticonvulsant; antidiabetic; gene therapy;
KW insulin dependent diabetes; multiple sclerosis; allergic condition;
KW lymphoid cell deficiency; asthma; myeloid cell deficiency; ulcer;
KW central nervous system disorder; osteoporosis; osteoarthritis;
KW Parkinson's disease; Huntington's disease; coagulation disorder;
KW haemophilia; thrombosis; inflammatory disorder; Crohn's disease;
KW tumour; infection; depression; psoriasis; ss.
XX
OS Homo sapiens.
XX
PN WO200021991-A1.
XX
PD 20-APR-2000.
XX
PF 15-OCT-1999; 99WO-US24206.
XX
PR 15-OCT-1998; 98US-0104436.
XX
PA (GENY) GENETICS INST INC.

XX Jacobus K, McCoy JM, Lavallie ER, Collins-Racie IA, Evans C;
 PI Merberg D, Treacy M, Bowman MR;
 XX WPI: 2000-317938/27.
 DR Isolated polynucleotides, and encoded proteins, comprising secreted
 PT expressed sequence tags (SESTs), useful for treating various disorders
 PT such as autoimmune, infectious, and central nervous system disorders -
 XX
 PS Claim 1: Page 723; 803pp; English.
 XX
 CC AAA43426 to AAA45925 represent specifically claimed secreted expressed
 CC sequence tags (SESTs), isolated from human, mouse, chicken and rat
 CC tissue sources. The SESTs can have a range of activities depending on
 CC the tissues they were isolated from. The activities include:
 CC chemotactic; proliferative; immunomodulatory; haematopoietic;
 CC chemokinetic; analgesic; haemostatic; thrombolytic; antiinflammatory;
 CC cyostatic; antibacterial; antifungal; antiviral; antidiabetic;
 CC antiallergic; antiparkinsonian; antipsoriatic; cerebroprotective;
 CC neurotrophic; antiparkinsonian; antipsoriatic; cerebroprotective;
 CC anticonvulsant; and antidepressant. The SESTs can be used for gene
 CC therapy and in vaccines. The SESTs are useful as probes for the
 CC identification and isolation of full-length cDNAs and genomic DNA
 CC molecules which correspond to the SESTs. Proteins encoded by the SESTs
 CC are useful in assays for determining biological activity and raising
 CC antibodies. They may be useful for treatment of autoimmune disorders
 CC (multiple sclerosis, insulin dependent diabetes), allergic conditions
 CC (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,
 CC osteoporosis, osteoarthritis, central nervous system disorders
 CC (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation
 CC disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's
 CC disease), tumours, bacterial, fungal or viral infections, depression and
 CC psoriasis. AAA45926 to AAA45931 represent linker variants which are given
 CC in the exemplification of the present invention.
 XX
 SO Sequence 286 BP; 44 A; 84 C; 69 G; 89 T; 0 other;

Query Match 52.0%; Score 13; DB 21; Length 286;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcaggtgcacgc 13
 |||||
 Db 93 GCAGGTGCATCC 81

RESULT 35
 AA214744/C
 ID AA214744 standard; cDNA: 300 BP.
 XX
 AC AA214744;
 XX
 DT 12-OCT-1999 (first entry)
 XX
 DE Human gene expression product cDNA sequence SEQ ID NO:2213.
 XX
 KW Human: gene: gene expression product; diagnosis: therapy; probe;
 KW detection; mapping; tissue typing; profiling; forensic; cancer;
 KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO9938972-A2.
 XX
 PD 05-AUG-1999.
 XX
 PF 28-JAN-1999; 99WO-US01619.
 XX
 PR 03-APR-1998; 98US-0080666.
 PR 28-JAN-1998; 98US-0072910.
 PR 24-FEB-1998; 98US-0075954.

PR 31-MAR-1998; 98US-0080114.
 PR 03-APR-1998; 98US-0080515.
 XX
 PA (CHIR) CHIRON CORP.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;
 PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;
 PI Jones WJ, Kassam A, Kennedy GC, Kita D, Labat I;
 PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
 PI Stache-Grain B, Sudduth-Klinger J, Williams LT;
 XX
 DR WPI: 1999-494092/41.
 XX
 PT Novel human genes and their expression products which are
 PT differentially expressed in different cell types
 XX
 PS Claim 1: Page 1159; 2479pp; English.
 XX
 CC The present invention describes a library of human polynucleotides
 CC comprising the sequences given in AA212532 to AA217779. Also described is
 CC a method of detecting differentially expressed genes correlated with the
 CC cancerous state of a mammalian cell, comprising detecting at least one
 CC differentially expressed gene product in a test sample from a cell
 CC suspected of being cancerous, where the gene product is encoded by one
 CC of the 5248 polynucleotide sequences given in AA212532 to AA217779. The
 CC polynucleotides can be used as a source of primers and probes, which can
 CC be used for a variety of purpose, e.g. detection of expression levels,
 CC mapping, tissue typing or profiling, forensics, genetic analysis and
 CC detection of polymorphisms. Polypeptides encoded by the polynucleotides
 CC can be used for raising antibodies for experimental, diagnostic and
 CC therapeutic purposes. The polynucleotides may also be used to construct
 CC arrays for diagnostics (which may be used to determine function of an
 CC encoded protein); and to detect differences in expression levels between
 CC two cells (e.g. to identify abnormal or diseased tissue in a human, to
 CC identify a genetic predisposition or susceptibility to a disease such as
 CC cancer). The polynucleotides of the invention are especially used in the
 CC diagnosis, prognosis and management of colorectal cancer, breast cancer,
 CC and lung cancer. The polynucleotides can also be used to screen for
 CC peptide analogues and antagonists.
 XX
 SO Sequence 300 BP; 83 A; 62 C; 70 G; 85 T; 0 other;

Query Match 52.0%; Score 13; DB 20; Length 300;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 gcacgcacacat 19
 |||||
 Db 174 GCATGCCACGAT 162

RESULT 36
 AA214726
 ID AA214726 standard; cDNA: 300 BP.
 XX
 AC AA214726;
 XX
 DT 12-OCT-1999 (first entry)
 XX
 DE Human gene expression product cDNA sequence SEQ ID NO:2195.
 XX
 KW Human: gene: gene expression product; diagnosis: therapy; probe;
 KW detection; mapping; tissue typing; profiling; forensic; cancer;
 KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO9938972-A2.
 XX
 PD 05-AUG-1999.
 XX

PF 28-JAN-1999; 99WO-US01619.
 XX
 PR 03-APR-1998; 98US-0080666.
 PR 28-JAN-1998; 98US-0072910.
 PR 24-FEB-1998; 98US-0075954.
 PR 31-MAR-1998; 98US-0080114.
 PR 03-APR-1998; 98US-0080315.
 XX
 PA (CHIR) CHIRON CORP.
 PA (HYSE-) HYSEQ INC.
 XX
 PI Crivenjakov R, Dickson M, Drmanac R, Drmanac S;
 PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;
 PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;
 PI Lanson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
 PI Stache-Crain B, Sudduth-Klinger J, Williams LT;
 DR WPI: 1999-494092/41.
 XX
 PT Novel human genes and their expression products which are
 PT differentially expressed in different cell types
 PS
 PS Claim 1; Page 1155; 2479pp; English.
 XX
 CC The present invention describes a library of human polynucleotides
 CC comprising the sequences given in AA12532 to AA17779. Also described is
 CC a method of detecting differentially expressed genes correlated with the
 CC cancerous state of a mammalian cell, comprising detecting at least one
 CC differentially expressed gene product in a test sample from a cell
 CC suspected of being cancerous, where the gene product is encoded by one
 CC of the 5248 polynucleotide sequences given in AA12532 to AA17779. The
 CC polynucleotides can be used as a source of primers and probes, which can
 CC be used for a variety of purpose, e.g. detection of expression levels,
 CC mapping, tissue typing or profiling, forensics, genetic analysis and
 CC detection of polymorphisms. Polypeptides encoded by the polynucleotides
 CC can be used for raising antibodies for experimental, diagnostic and
 CC therapeutic purposes. The polynucleotides may also be used to construct
 CC arrays for diagnostics (which may be used to determine function of an
 CC encoded protein); and to detect differences in expression levels between
 CC two cells (e.g. to identify abnormal or diseased tissue in a human, to
 CC identify a genetic predisposition or susceptibility to a disease such as
 CC cancer). The polynucleotides of the invention are especially used in the
 CC diagnosis, prognosis and management of colorectal cancer, breast cancer,
 CC and lung cancer. The polynucleotides can also be used to screen for
 CC peptide analogues and antagonists.
 CC
 SQ Sequence 300 BP; 80 A; 64 C; 63 G; 93 T; 0 other;

Query Match 52.0%; Score 13; DB 20; Length 300;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcaggtcgcgc 13
 |||||
 DB 180 gcaggtcgcgc 192

RESULT 37
 AAA31174/C
 ID AAA31174 standard; DNA; 322 BP.
 XX
 AC AAA31174;
 XX
 DT 05-JUL-2000 (first entry)
 XX
 DE Plant microsatellite marker #135.
 XX
 KW Plant microsatellite sequence; core repeat sequence; detection; probe;
 KW DNA polymorphism; genome mapping; physical mapping; fingerprinting;
 KW variety identification; genetic variability evaluation; primer; ss.
 XX
 OS Eucaelyptus grandis.

XX
 PN WO967421-A1.
 XX
 PD 29-DEC-1999.
 XX
 PF 25-JUN-1999; 99WO-NZ00092.
 XX
 PR 25-JUN-1998; 98US-0105307.
 XX
 PA (GENE-) GENESIS RES & DEV CORP LTD & FLETCHER.
 PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
 XX
 PI Havukkala JJ, Bloksberg LN, Glenn M;
 DR WPI: 2000-116958/10.
 XX
 PT New plant microsatellite markers and associated flanking species for
 PT the detection of polymorphic genetic markers -
 PS
 PS Claim 1; Page 116; 392pp; English.
 XX
 CC Sequences AAA31040-A32093 represent novel plant microsatellite sequences
 CC and associated flanking species. The sequences comprise a central core
 CC repeat sequence, especially selected from the sequences AAA32094-A32096
 CC with left and right flanking sequences. The polynucleotide sequences
 CC can be used in the detection of DNA polymorphisms, in genome mapping,
 CC in physical mapping, in positional cloning of genes, in variety
 CC identification and in evaluation of genetic variability within and
 CC between plant tissues, populations, cultivars, species and species
 CC groups. They may also be used to design hybridization probes for
 CC oligonucleotide fingerprinting and library screening and to design
 CC primers for microsatellite-primed PCR. Microsatellite markers are
 CC useful to locate specific economically useful genes in plant genomes.
 XX
 SQ Sequence 322 BP; 43 A; 101 C; 88 G; 89 T; 1 other;

Query Match 52.0%; Score 13; DB 21; Length 322;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 atcgccagcat 21
 |||||
 DB 21 atcgccagcat 9

RESULT 38
 AAT27962
 ID AAT27962 standard; DNA; 447 BP.
 XX
 AC AAT27962;
 XX
 DT 11-MAR-1997 (first entry)
 XX
 DE Hepatitis C virus type 10a isolate NN98 bases 478-925.
 XX
 KW Hepatitis C virus; subtype; polymerase chain reaction; amplification;
 KW PCR; primer; probe; antibody; infection; ss.
 XX
 OS Hepatitis C virus.
 XX
 PN WO9613590-A2.
 XX
 PD 09-MAY-1996.
 XX
 PF 23-OCT-1995; 95WO-EP04155.
 XX
 PR 28-JUN-1995; 95EP-0870076.
 PR 21-OCT-1994; 94EP-0870166.
 XX
 PA (INNO-) INNOGENETICS NV.
 XX
 PI Maertens G, Stuyver L;

XX WPI: 1996-251460/25.
 DR P-PSDB: AAR96551.
 XX Hepatitis C virus poly:nucleic acid unique to unidentified sub:type
 PT - used to develop probes and primers for new sub:types and vaccines
 PT to prevent and treat infection
 PS Claim 6; Fig 3; 150pp; English.
 XX The sequences AAT27937-T72989 represent novel sequences isolated from
 CC hepatitis C virus subtypes different from subtypes 1a-c, 2a-d, 3a-f,
 CC 4a-j, 5a and 6a. They esp. from the novel subtypes 1d-f, 2e-i, 2k, 2l,
 CC 3g, 4k-m, 7a-c or types 9, 10 or 11. The sequences corresp. to the 5'
 CC untranslated region (UR), the Core/E1, NS4 or NS5B regions of the HCV
 CC genome. This sequence represents nucleotides 478-925 from the HCV type
 CC 10a isolate NE98.
 CC The new HCV types were isolated from patients with chronic HCV from the
 CC Benelux countries, France, Cameroon and Vietnam, because of their
 CC aberrant reactivities. The RNA was extracted, cDNA synthesised and PCR
 CC amplified, cloned and genotyped. The 5'UR, Core/E1 and NS5B regions were
 CC sequenced either directly or partially and used to classify the new
 CC viruses into (sub)types based on comparison with known sequences.
 CC The sequences were used to generate the peptides AAR96424-R96524. The
 CC sequences can also be used to synthesise probes and primers for the
 CC detection of HCV in a sample. The polypeptides can be used to detect
 CC anti-HCV antibodies, for HCV typing or to prevent HCV infections.
 SO Sequence 447 BP; 82 A; 130 C; 114 G; 118 T; 3 other;

Query Match 52.0%; Score 13; DB 17; Length 447;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 catcgccagcatt 20
 |||||
 DB 406 catcgccagcatt 418

RESULT 39
 AAZ40788
 ID AAZ40788 standard; DNA; 538 BP.
 XX
 AC AAZ40788;

DT 18-JAN-2000 (first entry)
 XX

DE Secreted protein EST coding sequence 108-003-5-0-H2-FL.

XX Secreted protein; fingerprint identification technique;
 KW chromosome mapping; human; hereditary disease; diagnosis; cancer;
 KW hyperlipidaemia; cardiovascular; neurodegenerative disorder; therapy;
 KW autoimmune disease; rheumatic disease; embryogenic disorder; myopathy;
 KW renal injury; amino aciduria; hypoglycaemia; male rat infertility;
 KW hypertension; ss.

XX Homo sapiens.
 OS
 PN WO9940189-A2.
 PD 12-AUG-1999.
 PF 09-FEB-1999; 99WO-1B00282.
 PR 09-FEB-1998; 98US-0074121.
 PR 13-APR-1998; 98US-0081563.
 PR 10-APR-1998; 98US-0096116.
 PR 04-SEP-1998; 98US-0099273.
 XX (GEST) GENSET.
 PA Bougueleret L, Duclert A, Dumas Milne Edwards J;
 XX

XX WPI: 1999-600966/51.
 DR P-PSDB: AAF59660.
 XX Extended cDNAs useful for expressing secreted proteins and to obtain
 PT specific antibodies -
 PT
 PS Claim 1; Page 154-155; 244pp; English.

XX This sequence encodes a human secreted protein of the invention. The
 CC extended cDNAs (or genomic DNAs obtainable from them) may be used to
 CC prepare PCR primers and probes. These are useful for forensic matching or
 CC positive identification by DNA sequencing. They may also be used in
 CC alternative fingerprint identification techniques. Antibodies against the
 CC proteins encoded by the extended cDNAs are useful in identification of
 CC tissue types or cell species, as well as identifying tissue specific
 CC soluble proteins. The sequences can be used for chromosome mapping and
 CC identification of genes associated with hereditary diseases or drug
 CC response. Signal sequences from the cDNAs can be used in construction of
 CC secretion vectors. Other sequences derived from the extended cDNAs can be
 CC used to clone upstream genomic DNA sequences including promoters. This is
 CC in turn useful for identifying proteins that interact with promoter
 CC sequences. Some of the proteins may be useful in diagnosing and treating
 CC several disorders including, but not limited to: cancer, hyperlipidaemia,
 CC cardiovascular and neurodegenerative disorders, autoimmune diseases, and
 CC rheumatic diseases, embryogenic disorders, hypertension, renal injury,
 CC amino acidurias, hypoglycaemia, male rat infertility and myopathies.
 SO Sequence 538 BP; 104 A; 159 C; 151 G; 124 T; 0 other;

Query Match 52.0%; Score 13; DB 20; Length 538;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcaggtgcatgc 13
 |||||
 DB 314 gcaggtgcatgc 326

RESULT 40
 AAF09259
 ID AAF09259 standard; cDNA; 629 BP.
 XX
 AC AAF09259;

DT 13-MAR-2001 (first entry)
 XX

DE Fusarium venenatum EST SEQ ID NO:1782.

XX Multiple gene expression; filamentous fungal cell; EST;
 KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
 KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
 KW culture condition; environmental stress; spore morphogenesis;
 KW metabolic pathway engineering; catabolic pathway engineering; ss.

XX Fusarium venenatum.
 OS
 PN WO200056762-A2.
 PD 28-SEP-2000.
 PF 22-MAR-2000; 2000WO-US07781.
 PR 22-MAR-1999; 99US-0273623.
 XX (NOVO) NOVO NORDISK BIOTECH INC.
 PA (NOVO) NOVO NORDISK AS.
 XX Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
 XX WPI: 2000-594572/56.
 XX

PT Monitoring differential expression of genes in filamentous fungal cells
 PT uses fluorescence-labeled nucleic acids isolated from the cells and a
 PT substrate of expressed sequence tags -
 XX
 PS Claim 86; Page 1051; 3161pp; English.
 CC The present invention describes a method for monitoring differential
 CC expression of genes in a first filamentous fungal (FF) cell relative to
 CC expression of the same genes in one or more second filamentous fungal
 CC cells. The method uses fluorescence-labeled nucleic acids isolated from
 CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
 CC are used in the methods for monitoring differential expression of genes
 CC in a first filamentous fungal (FF) cell relative to expression of the
 CC same genes in one or more second filamentous fungal cells. Monitoring
 CC the global expression of genes from FF cells allows the production
 CC potential of the microorganisms to be improved. New genes may be
 CC discovered, possible functions of unknown open reading frames can be
 CC identified and gene copy number variation and stability can be
 CC monitored. The expression of genes can be used to study how FF cells
 CC adapt to changes in culture conditions, environmental stress, spore
 CC morphogenesis, recombination, metabolic or catabolic pathway
 CC engineering. Using ESTs provides several advantages over genomic or
 CC random cDNA clones including elimination of redundancy as one spot on an
 CC array equals one gene or open reading frame, and organisation of the
 CC microarrays based on function of the gene products to facilitate
 CC analysis of the results. AAF07478 to AAF11247 represents ESTs from
 CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from
 CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus
 CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
 CC all specifically claimed in the present invention.
 XX
 SQ Sequence 629 BP; 153 A; 169 C; 135 G; 168 T; 4 other;
 Query Match 52.0%; Score 13; DB 21; Length 629;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 8 catcgccagcatt 20
 |||||
 Db 83 catcgccagcatt 95
 RESULT 41
 AAF09268/C
 ID AAF09268 standard; cDNA; 632 BP.
 XX
 AC AAF09268;
 XX
 DT 13-MAR-2001 (first entry)
 XX
 DE Fusarium venenatum EST SEQ ID NO:1791.
 XX
 KM Multiple gene expression; filamentous fungal cell; EST;
 KM expressed sequence tag; Fusarium venenatum; Aspergillus niger;
 KM Aspergillus oryzae; Trichoderma reesei; identification; recombination;
 KM culture condition; environmental stress; spore morphogenesis;
 KM metabolic pathway engineering; catabolic pathway engineering; ss.
 OS
 XX Fusarium venenatum.
 XX
 PN WO200056762-A2.
 XX
 PD 28-SEP-2000.
 XX
 PF 22-MAR-2000; 2000WO-US07781.
 XX
 PR 22-MAR-1999; 99US-0273623.
 XX
 PA (NOVO) NOVO NORDISK BIOTECH INC.
 PA (NOVO) NOVO NORDISK AS.
 XX
 PI Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;

XX
 DR WPI: 2000-594572/56.
 XX
 PT Monitoring differential expression of genes in filamentous fungal cells
 PT uses fluorescence-labeled nucleic acids isolated from the cells and a
 PT substrate of expressed sequence tags -
 XX
 PS Claim 86; Page 1054; 3161pp; English.
 CC The present invention describes a method for monitoring differential
 CC expression of genes in a first filamentous fungal (FF) cell relative to
 CC expression of the same genes in one or more second filamentous fungal
 CC cells. The method uses fluorescence-labeled nucleic acids isolated from
 CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
 CC are used in the methods for monitoring differential expression of genes
 CC in a first filamentous fungal (FF) cell relative to expression of the
 CC same genes in one or more second filamentous fungal cells. Monitoring
 CC the global expression of genes from FF cells allows the production
 CC potential of the microorganisms to be improved. New genes may be
 CC discovered, possible functions of unknown open reading frames can be
 CC identified and gene copy number variation and stability can be
 CC monitored. The expression of genes can be used to study how FF cells
 CC adapt to changes in culture conditions, environmental stress, spore
 CC morphogenesis, recombination, metabolic or catabolic pathway
 CC engineering. Using ESTs provides several advantages over genomic or
 CC random cDNA clones including elimination of redundancy as one spot on an
 CC array equals one gene or open reading frame, and organisation of the
 CC microarrays based on function of the gene products to facilitate
 CC analysis of the results. AAF07478 to AAF11247 represents ESTs from
 CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from
 CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus
 CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
 CC all specifically claimed in the present invention.
 XX
 SQ Sequence 632 BP; 168 A; 168 C; 136 G; 148 T; 12 other;
 Query Match 52.0%; Score 13; DB 21; Length 632;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 4 ggtgcacgcagc 16
 |||||
 Db 547 GGTGCATCCGCCAG 535
 RESULT 42
 AAC55942/C
 ID AAC55942 standard; DNA; 646 BP.
 XX
 AC AAC55942;
 XX
 DT 25-JAN-2001 (first entry)
 XX
 DE Euclalyptus grandis transcription factor DNA sequence #73.
 XX
 KM Plant; transcription factor; gene expression; euclalyptus; pine; acacia;
 KM poplar; sweetgum; teak; mangrove; bZIP; G-box binding factor;
 KM basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;
 KM homeodomain zipper; LIM domain; AP2; ERBs; zinc finger domain;
 KM type 2 Cys2His2; CCAAT box element; MYB; ss.
 OS
 XX Euclalyptus grandis.
 XX
 PN WO200053724-A2.
 XX
 PD 14-SEP-2000.
 XX
 PF 09-MAR-2000; 2000WO-US06112.
 XX
 PR 11-MAR-1999; 99US-0266513.
 PR 18-AUG-1999; 99US-0149485.
 XX

PA (GENE-) GENESIS RES & DEV CORP LTD.
 PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
 XX
 PI Wood M, McGrath A, Shenk MA, Glenn M;
 XX WPI: 2000-579369/54.
 DR
 XX New isolated polynucleotide encoding a plant transcription factor for
 PT producing a plant e.g. a woody plant, preferably eucalyptus or pine,
 PT having modified gene expression or modified activity of a polypeptide
 PT
 XX
 PS Claim 1; Page 63; 747pp; English.
 XX
 CC The present invention relates to novel plant transcription factors from
 CC Eucalyptus grandis or Pinus radiata. The present sequence is the coding
 CC sequence for one such transcription factor. The transcription factor may
 CC be used to produce a plant having modified gene expression such as a
 CC woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or
 CC mahogany species or to modify the activity of a polypeptide in a plant.
 CC The transcription factors of the present invention are members from the
 CC following families of regulatory proteins: bZIP, bZIP family of G-box
 CC binding factors, basic helix-loop-helix zipper,
 CC homeo/c/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AP2
 CC and EREBS, zinc finger domains of type 2 Cys2His2, CCAAT box elements
 CC and MYB.
 XX
 SQ Sequence 646 BP; 129 A; 195 C; 174 G; 148 T; 0 other;
 Query Match 52.0%; Score 13; DB 21; Length 646;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 5 gtgcacgcacgc 17
 |||||||||
 DB 496 GTGCATCCCGCAGC 484
 RESULT 43
 AA217005
 ID AA217005 standard; cDNA: 714 BP.
 XX
 AC AA217005;
 XX
 DT 12-OCT-1999 (first entry)
 XX
 DE Human gene expression product cDNA sequence SEQ ID NO:4475.
 XX
 KW Human; gene; gene expression product; diagnosis; therapy; probe;
 KW detection; mapping; tissue typing; profiling; forensic; cancer;
 KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO9938972-A2.
 XX
 PD 05-AUG-1999.
 XX
 PF 28-JAN-1999; 99WO-US01619.
 XX
 PR 03-APR-1998; 98US-0080666.
 PR 28-JAN-1998; 98US-0072910.
 PR 24-FEB-1998; 98US-0075954.
 PR 31-MAR-1998; 98US-0080114.
 PR 03-APR-1998; 98US-0080515.
 XX
 PA (CHIR) CHIRON CORP.
 PA (HYSE-) HYSEQ INC.
 PI Ctkvenjakov R, Dickson M, Drmanac R, Drmanac S;
 PI Escobedo J, Garcia PD, Garcia V, Glese K, Innis MA;
 PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;

PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
 PI Stache-Crain B, Sudduth-Klinger J, Williams LT;
 XX
 XX WPI: 1999-494092/41.
 DR
 XX
 PT Novel human genes and their expression products which are
 PT differentially expressed in different cell types
 PS
 PS Claim 1; Page 2120; 2479pp; English.
 XX
 CC The present invention describes a library of human polynucleotides
 CC comprising the sequences given in AA212532 to AA21779. Also described is
 CC a method of detecting differentially expressed genes correlated with the
 CC cancerous state of a mammalian cell, comprising detecting at least one
 CC differentially expressed gene product in a test sample from a cell
 CC suspected of being cancerous, where the gene product is encoded by one
 CC of the 5248 polynucleotide sequences given in AA212532 to AA21779. The
 CC polynucleotides can be used as a source of primers and probes, which can
 CC be used for a variety of purpose, e.g. detection of expression levels,
 CC mapping, tissue typing or profiling, forensics, genetic analysis and
 CC detection of polymorphisms. Polypeptides encoded by the polynucleotides
 CC can be used for raising antibodies for experimental, diagnostic and
 CC therapeutic purposes. The polynucleotides may also be used to construct
 CC arrays for diagnostics (which may be used to determine function of an
 CC encoded protein); and to detect differences in expression levels between
 CC two cells (e.g. to identify abnormal or diseased tissue in a human, to
 CC identify a genetic predisposition or susceptibility to a disease such as
 CC cancer). The polynucleotides of the invention are especially used in the
 CC diagnosis, prognosis and management of colorectal cancer, breast cancer,
 CC and lung cancer. The polynucleotides can also be used to screen for
 CC peptide analogues and antagonists.
 XX
 SQ Sequence 714 BP; 186 A; 142 C; 146 G; 231 T; 9 other;
 Query Match 52.0%; Score 13; DB 20; Length 714;
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 tcgccagcattc 22
 |||||||||
 DB 74 tcgccagcattc 86
 RESULT 44
 AAF22436
 ID AAF22436 standard; cDNA: 725 BP.
 XX
 AC AAF22436;
 XX
 DT 26-MAR-2001 (first entry)
 XX
 DE Human breast cancer associated antigen nucleotide sequence SEQ ID NO:15.
 XX
 KW Human; breast cancer; gastric cancer; prostate cancer; diagnosis;
 KW cancer associated antigen; cytostatic; cancer vaccine; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200073801-A2.
 XX
 PD 07-DEC-2000.
 XX
 PF 26-MAY-2000; 2000WO-US14749.
 XX
 PR 28-MAY-1999; 99US-0136526.
 PR 10-SEP-1999; 99US-0153454.
 XX
 PA (LUDW-) LUDWIG INST CANCER RES.
 XX
 PI Obata Y;
 PI WPI: 2001-025274/03.

```

XX Nucleic acids encoding breast, gastric and prostate cancer associated
PT antigen precursors, useful for diagnosing and treating a condition
PT characterized by expression of an abnormal amount of a protein, e.g.
PT cancer
XX
XX Claim 50; Page 276; 799pp; English.
XX
CC AAF22422 to AAF22626, AAF22627 to AAF22773 and AAF22774 to AAF23014
CC represent nucleotide sequences encoding human breast, gastric and
CC prostate cancer associated antigen precursors (CAAP) respectively.
CC AAB63232 to AAB63467, AAB63468 to AAB63721 and AAB63722 to AAB63970
CC represent human breast, gastric and prostate CAAP protein sequence
CC respectively. CAAPs have cytostatic activity and can be used in the
CC production of cancer vaccines. The human CAAP proteins, peptides, nucleic
CC acids or anti-CAAP antibodies are useful for diagnosing and treating a
CC condition characterised by expression of an abnormal amount of a protein,
CC e.g. cancer.
XX
SQ Sequence 725 BP; 131 A; 225 C; 224 G; 136 T; 9 other;

Query Match
Best Local Similarity 52.0%; Score 13; DB 22; Length 725;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 gcatgccagcat 19
    |||||||||
Db 229 gcatgccagcat 241

RESULT 45
AAZ16470
ID AAZ16470 standard; cDNA; 749 BP.
XX
XX AAZ16470;
AC
XX
XX 12-OCt-1999 (first entry)
DT
XX
DE Human gene expression product cDNA sequence SEQ ID NO:3940.
XX
XX Human; gene: gene expression product; diagnosis; therapy; probe;
XX detection; mapping; tissue typing; profiling; forensic; cancer;
XX genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
XX
XX Homo sapiens.
OS
XX
XX WO9338972-A2.
PN
XX
XX 05-AUG-1999.
PD
XX
XX 28-JAN-1999; 99MO-US01619.
PF
XX
XX 03-APR-1998; 98US-0080666.
PR 28-JAN-1998; 98US-0072910.
PR 24-FEB-1998; 98US-0075954.
PR 31-MAR-1998; 98US-0080114.
PR 03-APR-1998; 98US-0080515.
XX
XX (CHIR ) CHIRON CORP.
PA
XX (HYSE-) HYSEQ INC.
XX
XX Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;
PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;
PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;
PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
PI Stache-Crain B, Sudduth-Klinger J, Williams LT;
XX
XX WPI; 1999-494092/41.
XX
XX Novel human genes and their expression products which are
XX differentially expressed in different cell types
XX

```

```

PS Claim 1; Page 1872; 2479pp; English.
XX
XX The present invention describes a library of human polynucleotides
CC comprising the sequences given in AAZ12532 to AAZ17779. Also described is
CC a method of detecting differentially expressed genes correlated with the
CC cancerous state of a mammalian cell, comprising detecting at least one
CC differentially expressed gene product in a test sample from a cell
CC suspected of being cancerous, where the gene product is encoded by one
CC of the 5248 polynucleotide sequences given in AAZ12532 to AAZ17779. The
CC polynucleotides can be used as a source of primers and probes, which can
CC be used for a variety of purpose, e.g. detection of expression levels,
CC mapping, tissue typing or profiling, forensics, genetic analysis and
CC detection of polymorphisms. Polypeptides encoded by the polynucleotides
CC can be used for raising antibodies for experimental, diagnostic and
CC therapeutic purposes. The polynucleotides may also be used to construct
CC arrays for diagnostics (which may be used to determine function of an
CC encoded protein); and to detect differences in expression levels between
CC two cells (e.g. to identify abnormal or diseased tissue in a human, to
CC identify a genetic predisposition or susceptibility to a disease such as
CC cancer). The polynucleotides of the invention are especially used in the
CC diagnosis, prognosis and management of colorectal cancer, breast cancer,
CC and lung cancer. The polynucleotides can also be used to screen for
CC peptide analogues and antagonists.
XX
SQ Sequence 749 BP; 182 A; 138 C; 154 G; 258 T; 17 other;

Query Match
Best Local Similarity 52.0%; Score 13; DB 20; Length 749;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gccagtgcatgc 13
    |||||||||
Db 260 gccagtgcatgc 272

```

Search completed: October 9, 2001, 15:52:31
Job time: 15112 sec

Wed Oct 10 07:46:00 2001

us-09-396-196f-6.oli.rml

Page 1

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 15:55:32 ; Search time 218.82 Seconds
(without alignments)
21.629 Million cell updates/sec

Title: US-09-396-196f-6

Perfect score: 25

Sequence: 1 gcaggtgcacgcacgattcgcgt 25

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 324599 seqs, 94655562 residues

Word size: 9

Total number of hits satisfying chosen parameters: 9730

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database:

Issued Patents NA:
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS.COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfile1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	25	100.0	1041	2	US-08-401-068-7
2	25	100.0	1041	2	US-08-846-338-7
3	25	100.0	5872	3	US-08-411-768B-1
4	25	100.0	5872	3	US-08-411-768B-6
5	14	56.0	718	4	US-09-142-078-45
6	14	56.0	2193	4	US-08-420-235B-16
7	14	56.0	2193	4	US-08-793-624-16
8	14	56.0	2193	5	PCT-US95-10194-16
9	14	56.0	2232	5	PCT-US96-05320A-113
10	14	56.0	2973	2	US-08-343-101A-5
11	14	56.0	2973	2	US-09-183-688-5
12	14	56.0	5319	1	US-08-169-927-1
13	14	56.0	20710	1	US-08-420-235B-1
14	14	56.0	20710	1	US-08-793-624-1
15	14	56.0	20710	5	PCT-US95-10194-1
16	14	56.0	35100	2	US-08-770-379-18
17	14	56.0	35100	4	US-08-757-669A-18
18	13	52.0	304	4	US-09-060-756-726
19	13	52.0	343	4	US-09-060-756-726
20	13	52.0	447	4	US-08-836-075A-51
21	13	52.0	1460	2	US-08-933-750C-60
22	13	52.0	1460	2	US-09-234-613-60
23	13	52.0	1624	2	US-07-862-588B-6
24	13	52.0	1775	2	US-07-862-588B-5
25	13	52.0	2290	1	US-08-584-226-1
26	13	52.0	2312	2	US-09-134-566-15
27	13	52.0	2314	2	US-09-134-566-8

C 28	13	52.0	8878	1	US-08-759-444-2	Sequence 2, Appl
C 29	13	52.0	9880	3	US-08-680-897-1	Sequence 1, Appl
C 30	12	48.0	19	3	US-08-851-843A-92	Sequence 92, Appl
C 31	12	48.0	19	4	US-08-974-549A-984	Sequence 384, Appl
C 32	12	48.0	38	1	US-08-373-124A-1692	Sequence 1692, Ap
C 33	12	48.0	38	1	US-08-435-628-1692	Sequence 59, Appl
C 34	12	48.0	61	3	US-09-023-228B-59	Sequence 16, Appl
C 35	12	48.0	69	2	US-08-484-552A-16	Sequence 16, Appl
C 36	12	48.0	69	2	US-08-484-552A-16	Sequence 16, Appl
C 37	12	48.0	69	5	PCT-US96-09472-16	Sequence 93, Appl
C 38	12	48.0	386	4	US-08-998-416-93	Sequence 184, Appl
C 39	12	48.0	438	2	US-08-743-637B-184	Sequence 56, Appl
C 40	12	48.0	656	4	US-08-998-416-56	Sequence 175, Appl
C 41	12	48.0	660	2	US-08-743-637B-175	Sequence 175, Appl
C 42	12	48.0	660	3	US-08-526-840B-175	Sequence 175, Appl
C 43	12	48.0	715	4	US-08-998-416-767	Sequence 237, Appl
C 44	12	48.0	754	4	US-08-998-416-237	Sequence 39, Appl
C 45	12	48.0	841	4	US-08-990-823-39	

ALIGNMENTS

RESULT 1
US-08-401-068-7
Sequence 7, Application US/08401068
Patent No. 5859335
GENERAL INFORMATION:
APPLICANT: Patton, David
TITLE OF INVENTION: Enhanced Biotin Biosynthesis in Plant Tissue
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESSES:
ADDRESS: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/401,068
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/351,970
FILING DATE: 08-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Fimer, James Scott
REGISTRATION NUMBER: 36,129
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8614
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1041 base pairs
TYPE: nucleic acid
STRANDEDNESS: Single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1038
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /product= "Biotin synthase"
OTHER INFORMATION: /evidence= EXPERIMENTAL
US-08-401-068-7

Query Match 100.0%; Score 25; DB 2; Length 1041;
Best Local Similarity 100.0%; Pred. No. 6.5e-06;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gcaggtgcatcgccagcatttcgat 25
|||||
DB 84 GCAGGTGTCATCGCCAGCATTTCGAT 108

RESULT 2
US-08-846-338-7
; Sequence 7, Application US/08846338
; Patent No. 5869719
; GENERAL INFORMATION:
; APPLICANT: Patton, David
; TITLE OF INVENTION: Transgenic Plants Having Increased Biotin Content
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: No. 5869719artis Corporation
; STREET: 520 White Plains Road, P.O. Box 2005
; CITY: Tarrytown
; STATE: NY
; COUNTRY: USA
; ZIP: 10591
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/846,338
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8587
; TELEFAX: 919-541-8589
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1041 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEetical: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1038
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /product="biotin synthase"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
US-08-846-338-7

Query Match 100.0%; Score 25; DB 2; Length 1041;
Best Local Similarity 100.0%; Pred. No. 6.5e-06;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gcaggtgcatcgccagcatttcgat 25
|||||
DB 84 GCAGGTGTCATCGCCAGCATTTCGAT 108

RESULT 3
US-08-411-768B-1
; Sequence 1, Application US/08411768B
; Patent No. 6083712
; GENERAL INFORMATION:
; APPLICANT: Olwen Birch
; APPLICANT: Johann Brass
; APPLICANT: Martin Fuhrmann

APPLICANT: Nicholas Shaw
; TITLE OF INVENTION: Biotechnological Method
; TITLE OF INVENTION: of Producing Biotin
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Wordperfect
; SOFTWARE: Version 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/411,768B
; FILING DATE: 31-March-95
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 3124/92
; FILING DATE: 02-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 2134/93
; FILING DATE: 15-JUL-1993
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5872 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEtical: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli
; STRAIN: DSM498
; IMMEDIATE SOURCE:
; CLONE: pBO30A-15/9
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 117..1157
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /codon_start= 117
; OTHER INFORMATION: /product= "Biotin synthase"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /gene= "biob"
; OTHER INFORMATION: /number= 1
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2295..3050
; OTHER INFORMATION: /codon_start= 2295
; OTHER INFORMATION: /function= "involved in pimeloyl-CoA synthesis"
; OTHER INFORMATION: /product= "protein"
; OTHER INFORMATION: /gene= "bioc"
; OTHER INFORMATION: /number= 3
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 3750..5039
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /codon_start= 3750
; OTHER INFORMATION: /EC_number= 2.6.1.62
; OTHER INFORMATION: /product= "DAPA synthase"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /gene= "bioA"
; OTHER INFORMATION: /number= 5
; OTHER INFORMATION: /standard_name= "S-Adenosyl-L-methionine:8-amino-7-oxononanoate
; OTHER INFORMATION: /aminotransf."

LOCATION: 5098..5574
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start= 5098
OTHER INFORMATION: /function= "unknown", involved in biotin synthesis"
OTHER INFORMATION: /product= "protein"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "ORF1"
OTHER INFORMATION: /number= 6
FEATURE:
NAME/KEY: -10_signal
LOCATION: 45..49
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /standard_name= "promoter plac"
FEATURE:
NAME/KEY: -35_signal
LOCATION: 23..28
OTHER INFORMATION: /standard_name= "promoter plac"
FEATURE:
NAME/KEY: RBS
LOCATION: 105..119
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /standard_name= "biob RBS no.9"
FEATURE:
NAME/KEY: RBS
LOCATION: 2284..2297
OTHER INFORMATION: /standard_name= "bioc RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 3742..3752
OTHER INFORMATION: /standard_name= "bioa RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 5088..5100
OTHER INFORMATION: /standard_name= "ORF1 RBS"
FEATURE:
NAME/KEY: terminator
LOCATION: 5583..5644
OTHER INFORMATION: /standard_name= "rho-independent
OTHER INFORMATION: transcriptional terminator"
FEATURE:
NAME/KEY: stem_loop
LOCATION: 5583..5605
FEATURE:
NAME/KEY: promoter
LOCATION: 1..96
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /function= "promoter plac"
OTHER INFORMATION: /evidence= EXPERIMENTAL
PUBLICATION INFORMATION:
DOCUMENT NUMBER: WO 87/01391 B1
FILING DATE: 26-AUG-1986
PUBLICATION DATE: 07-APR-1993
US-08-411-768B-1
Query Match 100.0%; Score 25; DB 3; Length 5872;
Best local Similarity 100.0%; Pred. No. 6.7e-06;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gcaggtgcattcgccagcattcgat 25
DB 200 GCAGGTGCATCGCCAGCATTTCGAT 224
US-08-411-768B-6
; Sequence 6, Application US/08411768B
; Patent No. 6083712
; GENERAL INFORMATION:
; APPLICANT: Olwen Birch
; APPLICANT: Johann Brass

APPLICANT: Martin Fuhrmann
APPLICANT: Nicholas Shaw
TITLE OF INVENTION: Biotechnological Method
TITLE OF INVENTION: of Producing Biotin
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Brunbaugh, Graves, Donohue & Raymond
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/411,768B
FILING DATE: 31-March-95
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 3124/92
FILING DATE: 02-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 2134/93
FILING DATE: 15-JUL-1993
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 5872 base pairs
TYPE: nucleic acid
STRANDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYDROTHERMAL: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli
STRAIN: DSM498
IMMEDIATE SOURCE:
CLONE: pBO30A15-9
FEATURE:
NAME/KEY: CDS
LOCATION: 1154..2308
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start= 1154
OTHER INFORMATION: /EC_number= 2.3.1.47
OTHER INFORMATION: /product= "KAPA synthase"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "bioF"
OTHER INFORMATION: /number= 2
OTHER INFORMATION: /standard_name= "8-Amino-7-oxononanoate synthase"
FEATURE:
NAME/KEY: CDS
LOCATION: 3043..3753
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /codon_start= 3043
OTHER INFORMATION: /EC_number= 6.3.3.3
OTHER INFORMATION: /product= "DTB synthase"
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /gene= "biob"
OTHER INFORMATION: /number= 4
OTHER INFORMATION: /standard_name= "Dethiobiotin synthase"
FEATURE:
NAME/KEY: RBS
LOCATION: 1141..1156
OTHER INFORMATION: /standard_name= "biob RBS"
FEATURE:
NAME/KEY: RBS
LOCATION: 3030..3045
OTHER INFORMATION: /standard_name= "biob RBS"
PUBLICATION INFORMATION:
DOCUMENT NUMBER: WO 87/01391 B1

FILED DATE: 26-AUG-1986
PUBLICATION DATE: 07-APR-1993
US-08-411-768B-6

Query Match 100.0%; Score 25; DB 3; Length 5872;
Best Local Similarity 100.0%; Pred. No. 6,7e-06;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcaggtgcacgcacgcatcgcacat 25
|||||
DB 200 GCAGGTGCATGCCAGCATTCGAT 224

RESULT 5
US-09-142-078-45/C
Sequence 45, Application US/09142078
Patent No. 6172041
GENERAL INFORMATION:
APPLICANT: McCabe, R. Tyler
APPLICANT: Zhou, Li-Ming
APPLICANT: Layer, Richard T.
TITLE OF INVENTION: Use of Conantokins
NUMBER OF SEQUENCES: 71
CORRESPONDENCE ADDRESS:
ADDRESSEE: Rothwell, Figg, Ernst & Kurtz, P.C.
STREET: 555 Thirteenth Street, N.W., Suite 701-E
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/142,078
FILING DATE: 10-FEB-1999
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO US97/12652
FILING DATE: 21-JUL-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/762,377
FILING DATE: 06-DEC-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/684,750
FILING DATE: 22-JUL-1996
ATTORNEY/AGENT INFORMATION:
NAME: Innen, Jeffrey L.
REGISTRATION NUMBER: 28,957
REFERENCE/DOCKET NUMBER: 2314-135, A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-783-6040
TELEFAX: 202-783-6031
INFORMATION FOR SEQ ID NO: 45:
SEQUENCE CHARACTERISTICS:
LENGTH: 718 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (cDNA)
ORIGINAL SOURCE:
ORGANISM: Conus geographus
FEATURE:
NAME/KEY: CDS
LOCATION: 110..409
US-09-142-078-45

Query Match 56.0%; Score 14; DB 4; Length 718;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 caggtgcacgcacca 15
|||||
DB 118 CAGGTGCATGCCCA 105

RESULT 6
US-08-420-235B-16/C
Sequence 16, Application US/08420235B
Patent No. 5801042
GENERAL INFORMATION:
APPLICANT: Chang, Yuan
APPLICANT: Moore, Patrick S.
TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
NUMBER OF SEQUENCES: 47
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/420,235B
FILING DATE:
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 45185-B
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 2193 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: N
ANTI-SENSE: N
FEATURE:
NAME/KEY: CDS
LOCATION: 1..2193
OTHER INFORMATION:
US-08-420-235B-16

Query Match 56.0%; Score 14; DB 1; Length 2193;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 tgcacgcacgacat 19
|||||
DB 46 TGCATGCCAGCAT 33

RESULT 7
US-08-793-624-16/C
Sequence 16, Application US/08793624C
Patent No. 6150093
GENERAL INFORMATION:
APPLICANT: Chang, Yuan
APPLICANT: Moore, Patrick S.
TITLE OF INVENTION: Unique Associated Kaposi's Sarcoma Virus Sequences And

FILE REFERENCE: 45185-C-PCT-US/JPW
CURRENT APPLICATION NUMBER: US/08/793,624C
CURRENT FILING DATE: 1997-02-18
NUMBER OF SEQ ID NOS: 58
SOFTWARE: Patentl Ver. 2.1
SEQ ID NO 16
LENGTH: 2193
TYPE: DNA
ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-08-793-624-16

Query Match 56.0%; Score 14; DB 3; Length 2193;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 tgcacgcacgat 19
|||||
Db 46 TGCATGCCAGCAT 33

RESULT 8

PCT-US95-10194-16/c
Sequence 16, Application PC/TUS9510194
GENERAL INFORMATION:
APPLICANT: The Trustees of Columbia University in the City of New York
APPLICANT: City
TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
TITLE OF INVENTION: SEQUENCES AND USES THEREOF
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentl Release #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10194
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 45185-C-PCT/JPW/MS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 2193 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: N
ANTI-SENSE: N
FEATURE:
NAME/KEY: CDS
LOCATION: 1..2193
OTHER INFORMATION:
PCT-US95-10194-16

Query Match 56.0%; Score 14; DB 5; Length 2193;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 tgcacgcacgat 19
|||||
Db 46 TGCATGCCAGCAT 33

RESULT 9

PCT-US96-05320A-113/c
Sequence 113, Application PC/TUS9605320A
GENERAL INFORMATION:
APPLICANT: Human Genome Sciences
APPLICANT: 9410 Key West Avenue
APPLICANT: Rockville, MD 20850
APPLICANT: United States of America
APPLICANT: Johns Hopkins University
APPLICANT: 720 Rutland Avenue
APPLICANT: Baltimore, MD 21205
APPLICANT: United States of America
APPLICANT: Mark D. Adams
APPLICANT: Owen White
APPLICANT: Hamilton O. Smith
APPLICANT: J. Craig Venter
TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus Influenzae Rd Genome
NUMBER OF SEQUENCES: 48
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1100 New York Avenue, Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20003-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44MB storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/05320A
FILING DATE: April 22, 1996
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/476,102
FILING DATE: June 7, 1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/487,429
FILING DATE: June 7, 1995
ATTORNEY/AGENT INFORMATION:
NAME: Eric K. Steffe
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.014PC01
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 113:
SEQUENCE CHARACTERISTICS:
LENGTH: 2232 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
PCT-US96-05320A-113

Query Match 56.0%; Score 14; DB 5; Length 2232;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 9 atgcacgatc 22
|||||
Db 636 ATGCCAGCATTTG 623

RESULT 10
US-08-343-101A-5/c
Sequence 5, Application US/08343101A

Patent No. 5830759
GENERAL INFORMATION:
APPLICANT: Chang, Yuan
APPLICANT: Moore, Patrick S.
TITLE OF INVENTION: Unique Associated Kaposi's Sarcoma
TITLE OF INVENTION: Virus Sequences And Uses Thereof
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/343,101A
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: White Esq., John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 45185-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-278-0400
TELEFAX: 212-391-0526
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2973 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: N
ANTI-SENSE: N
US-08-343-101A-5

Query Match 56.0%; Score 14; DB 2; Length 2973;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 6 tgcattgccagcat 19
|||||
DB 2182 TGCATGCCAGCAT 2169

RESULT 11
US-09-183-688-5/c
Sequence 5, Application US/09183688
Patent No. 6093550
GENERAL INFORMATION:
APPLICANT: Chang, Yuan
APPLICANT: Moore, Patrick S.
TITLE OF INVENTION: Unique Associated Kaposi's Sarcoma
TITLE OF INVENTION: Virus Sequences And Uses Thereof
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.24

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/183,688
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/343,101
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: White Esq., John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 45185-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-278-0400
TELEFAX: 212-391-0526
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2973 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: N
ANTI-SENSE: N
US-09-183-688-5

Query Match 56.0%; Score 14; DB 3; Length 2973;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 6 tgcattgccagcat 19
|||||
DB 2182 TGCATGCCAGCAT 2169

RESULT 12
US-08-169-927-1/c
Sequence 1, Application US/08169927
Patent No. 5783441
GENERAL INFORMATION:
APPLICANT: Carl, Mitchell
APPLICANT: Dobson, Michael E.
APPLICANT: Ching, Wei Mei
APPLICANT: Dasch, Gregory A
TITLE OF INVENTION: Gene and Protein Applicable to the
TITLE OF INVENTION: Preparation of Vaccines for Rickettsia typhi and the Detection of Both
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Counsel, Naval Medical R & D Command
STREET: Bldg. 1, T-12, 8901 Wisconsin Ave.
CITY: Bethesda
STATE: MD
COUNTRY: USA
ZIP: 20889-5606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/169,927
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/742,128
FILING DATE: 08/09/91
ATTORNEY/AGENT INFORMATION:
NAME: Spevack, A. David
REGISTRATION NUMBER: 24,743
REFERENCE/DOCKET NUMBER: 75,976
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 295-6759

```

; TELEFAX: (301) 295-1022
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5319 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; ORGANISM: Rickettsia prowazekii
; STRAIN: Breinl
; FEATURE:
; NAME/KEY: -35_signal
; LOCATION: 340..345
; FEATURE:
; NAME/KEY: -10_signal
; LOCATION: 363..368
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 391..5226
; FEATURE:
; NAME/KEY: RBS
; LOCATION: 379..386
; FEATURE:
; NAME/KEY: stem_loop
; LOCATION: 5270..5306
; PUBLICATION INFORMATION:
; AUTHORS: Carl, M.
; AUTHORS: Dobson, M. E.
; AUTHORS: Ching, W. M.
; AUTHORS: Dasch, G. A.
; TITLE: Characterization of the gene encoding the
; TITLE: protective S-layer protein of Rickettsia
; TITLE: prowazekii, presence of a truncated identical
; TITLE: homolog in Rickettsia typhi
; JOURNAL: Proc. Natl. Acad. Sci. U.S.A.
; DATE: 1990
; RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 5319
US-08-169-927-1

Query Match 56.0%; Score 14; DB 1; Length 5319;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 catgcgcagcat 21
|||||
DB 2221 CATGCCAGCATTT 2208

RESULT 13
US-08-420-235B-1/c
; Sequence 1, Application US/08420235B
; Patent No. 5801042
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
; TITLE OF INVENTION: SEQUENCES AND USES THEREOF
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/420,235B
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 45185-B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20710 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: N
; ANTI-SENSE: N
US-08-420-235B-1

Query Match 56.0%; Score 14; DB 1; Length 20710;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 tgcacgcgcagcat 19
|||||
DB 2138 TGCATCGCCAGCAT 2125

```

```

RESULT 14
US-08-793-624-1/c
; Sequence 1, Application US/08793624C
; Patent No. 6150093
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: Unique Associated Kaposi's Sarcoma Virus Sequences and
; TITLE OF INVENTION: Uses Thereof
; FILE REFERENCE: 45185-C-PCT-US/JPW
; CURRENT APPLICATION NUMBER: US/08/793,624C
; CURRENT FILING DATE: 1997-02-18
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 20710
; TYPE: DNA
; ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-08-793-624-1

Query Match 56.0%; Score 14; DB 3; Length 20710;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 tgcacgcgcagcat 19
|||||
DB 2138 TGCATCGCCAGCAT 2125

RESULT 15
PCT-US95-10194-1/c
; Sequence 1, Application PC/TUS9510194
; GENERAL INFORMATION:
; APPLICANT: The Trustees of Columbia University in the City of New York
; APPLICANT: City
; TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
; TITLE OF INVENTION: SEQUENCES AND USES THEREOF
; NUMBER OF SEQUENCES: 45

```

;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Cooper & Dunham LLP
;; STREET: 1185 Avenue of the Americas
;; CITY: New York
;; STATE: New York
;; COUNTRY: U.S.A.
;; ZIP: 10036
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: PCT/US95/10194
;; FILING DATE:
;; CLASSIFICATION:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: White, John P.
;; REGISTRATION NUMBER: 28,678
;; REFERENCE/DOCKET NUMBER: 45185-C-PCT/JPW/MSC
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (212) 278-0400
;; TELEFAX: (212) 391-0525
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 20710 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: DNA (genomic)
;; HYPOTHEetical: N
;; ANTI-SENSE: N
;; PCT-US95-10194-1

Query Match 56.0%; Score 14; DB 5; Length 20710;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 tgcacgcacgacat 19
|||||

Db 2138 TGCATGCCGACAT 2125

RESULT 16
US-08-770-379-18/c
; Sequence 18, Application US/08770379
; Patent No. 5849564
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A.
; APPLICANT: Russo, James J.
; APPLICANT: Edelman, Isidore S.
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/770,379
; FILING DATE:
; CLASSIFICATION: 435

;; ATTORNEY/AGENT INFORMATION:
;; NAME: White, John P.
;; REGISTRATION NUMBER: 28,678
;; REFERENCE/DOCKET NUMBER: 52342
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (212) 278-0400
;; TELEFAX: (212) 391-0525
;; INFORMATION FOR SEQ ID NO: 18:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 35100 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; MOLECULE TYPE: DNA (genomic)
;; US-08-770-379-18

Query Match 56.0%; Score 14; DB 2; Length 35100;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 tgcacgcacgacat 19
|||||

Db 2058 TGCATGCCGACAT 2045

RESULT 17
US-08-757-669A-18/c
; Sequence 18, Application US/08757669A
; Patent No. 6183751
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A.
; APPLICANT: Russo, James J.
; APPLICANT: Edelman, Isidore S.
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/757,669A
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 45185-F
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 35100 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-757-669A-18

Query Match 56.0%; Score 14; DB 4; Length 35100;

Best Local Similarity 100.0%; Pred. No. 12;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 gtcatgccagcatt 19
|||||
Db 2058 TGCAATGCCAGCATT 2045

RESULT 18

US-09-060-756-726
; Sequence 726, Application US/09060756
; Patent No. 6183957
; GENERAL INFORMATION:
; APPLICANT: Cole, Stewart
; APPLICANT: Buchrieser-Brosch, Roland
; APPLICANT: Gordon, Stephen
; APPLICANT: Billault, Alain
; TITLE OF INVENTION: METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST FROM
; TITLE OF INVENTION: THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED DNA
; TITLE OF INVENTION: LIBRARY APPLICATION TO THE DETECTION OF MYCOBACTERIA
; FILE REFERENCE: 3495-0169
; CURRENT APPLICATION NUMBER: US/09/060,756
; CURRENT FILING DATE: 1998-04-16
; NUMBER OF SEQ ID NOS: 743
; SOFTWARE: Patentlin Ver. 2.0
; SEQ ID NO 726
; LENGTH: 304
; TYPE: DNA
; ORGANISM: Mycobacterium bovis
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (various positions within the sequence)
; OTHER INFORMATION: applicants are uncertain of bases designated as "n"
US-09-060-756-726

Query Match 52.0%; Score 13; DB 4; Length 304;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 gtcatgccagcatt 17
|||||
Db 136 gtcatgccagcatt 148

RESULT 19

US-09-060-756-570
; Sequence 570, Application US/09060756
; Patent No. 6183957
; GENERAL INFORMATION:
; APPLICANT: Cole, Stewart
; APPLICANT: Buchrieser-Brosch, Roland
; APPLICANT: Gordon, Stephen
; APPLICANT: Billault, Alain
; TITLE OF INVENTION: METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST FROM
; TITLE OF INVENTION: THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED DNA
; TITLE OF INVENTION: LIBRARY APPLICATION TO THE DETECTION OF MYCOBACTERIA
; FILE REFERENCE: 3495-0169
; CURRENT APPLICATION NUMBER: US/09/060,756
; CURRENT FILING DATE: 1998-04-16
; NUMBER OF SEQ ID NOS: 743
; SOFTWARE: Patentlin Ver. 2.0
; SEQ ID NO 570
; LENGTH: 343
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
US-09-060-756-570

Query Match 52.0%; Score 13; DB 4; Length 343;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 gtcatgccagcatt 17
|||||
Db 135 gtcatgccagcatt 147

RESULT 20

US-08-836-075A-51
; Sequence 51, Application US/08836075A
; Patent No. 6180768
; GENERAL INFORMATION:
; APPLICANT: MAERTENS, GEERT
; APPLICANT: STUYVER, LIEVEN
; TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
; TITLE OF INVENTION: AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC
; NUMBER OF SEQUENCES: 207
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ARNOLD, WHITE & DURKEE
; STREET: P.O. BOX 4433
; CITY: HOUSTON
; STATE: TEXAS
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Microsoft Word 6.0 / ASCII text output
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/836,075A
; FILING DATE: 21 Apr 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP95/04155
; FILING DATE: 23 Oct 1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 94870166.9
; FILING DATE: 21 Oct 1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 95870076.7
; FILING DATE: 28 Jun 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: KAMMERER, PATRICIA A.
; REGISTRATION NUMBER: 29,775
; REFERENCE/DOCKET NUMBER: INNS:004
; INFORMATION FOR SEQ ID NO: 51:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 447 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
US-08-836-075A-51

Query Match 52.0%; Score 13; DB 4; Length 447;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 catgccagcatt 20
|||||
Db 406 CATGCCAGCATT 418

RESULT 21

US-08-933-750C-60
; Sequence 60, Application US/08933750C
; Patent No. 5932442
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Bandman, Olga

APPLICANT: Shah, Purvi
APPLICANT: Au-Young, Janice
APPLICANT: Yue, Henry
APPLICANT: Guegler, Karl J.
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
NUMBER OF SEQUENCES: 98
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/933 750C
FILING DATE: September 23, 1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0356 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 60:
SEQUENCE CHARACTERISTICS:
LENGTH: 1460 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: T1YMNOR01
CLONE: 140516
US-08-933-750C-60

Query Match 52.0%; Score 13; DB 2; Length 1460;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 ccagcattcgat 25
DB 722 CCAGCATTTCGAT 734

RESULT 22
US-09-234-613-60
Sequence 60, Application US/09234613
Patent No. 6132973
GENERAL INFORMATION:
APPLICANT: Lal, Preeti
APPLICANT: Hillman, Jennifer L.
APPLICANT: Bandman, Olga
APPLICANT: Shah, Purvi
APPLICANT: Au-Young, Janice
APPLICANT: Yue, Henry
APPLICANT: Guegler, Karl J.
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
NUMBER OF SEQUENCES: 98
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive

CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/234 613
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/933 750
FILING DATE: September 23, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0356 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 60:
SEQUENCE CHARACTERISTICS:
LENGTH: 1460 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: T1YMNOR01
CLONE: 140516
US-09-234-613-60

Query Match 52.0%; Score 13; DB 3; Length 1460;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 ccagcattcgat 25
DB 722 CCAGCATTTCGAT 734

RESULT 23
US-07-862-588B-6
Sequence 6, Application US/07862588B
Patent No. 5916796
GENERAL INFORMATION:
APPLICANT: Joergensen, Per Linna
APPLICANT: Sch Ieln, Martin
APPLICANT: Hansen, Christian
TITLE OF INVENTION: An Enzyme Exhibiting Cellulase Activity
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5916796 No. 5916796disk of No. 5916796th America, Inc.
STREET: 405 Lexington Avenue, 62nd floor
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10017
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/862,588B
FILING DATE: 19920727
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DK 164/90

FILING DATE: 19-JAN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/DK91/00013
FILING DATE: 18-JAN-1991
ATTORNEY/AGENT INFORMATION:
NAME: Zelson, Steve T. / Lambdiris, Elias J.
REGISTRATION NUMBER: 30,335 / 33,728
REFERENCE/DOCKET NUMBER: 3425.204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 867 0123
TELEFAX: 212 867 0298
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1624 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: Bacillus lautus
STRAIN: NCIMB 40250
FEATURE:
NAME/KEY: CDS
LOCATION: 30..1607
OTHER INFORMATION:
US-07-862-588B-6

Query Match 52.0%; Score 13; DB 2; Length 1624;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 caggtgcacgcc 14
|||||
Db 934 CAGGTGCATGCC 946

RESULT 24
US-07-862-588B-5
Sequence 5, Application US/07862588B
Patent No. 5916796
GENERAL INFORMATION:
APPLICANT: Joergensen, Per Linna
APPLICANT: Sch lein, Martin
APPLICANT: Hansen, Christian
TITLE OF INVENTION: An Enzyme exhibiting Cellulase Activity
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 59167960 No. 5916796disk of No. 5916796th America, Inc.
STREET: 405 Lexington Avenue, 62nd floor
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10017
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/862,588B
FILING DATE: 19920727
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DK 164/90
FILING DATE: 19-JAN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/DK91/00013
FILING DATE: 18-JAN-1991
ATTORNEY/AGENT INFORMATION:
NAME: Zelson, Steve T. / Lambdiris, Elias J.
REGISTRATION NUMBER: 30,335 / 33,728

REFERENCE/DOCKET NUMBER: 3425.204-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 867 0123
TELEFAX: 212 867 0298
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1775 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: Bacillus lautus
STRAIN: NCIMB 40250
FEATURE:
NAME/KEY: CDS
LOCATION: 30..(1625,1775)
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /partial
OTHER INFORMATION: /evidence= EXPERIMENTAL
OTHER INFORMATION: /transl_except= (pos: 1446 .. 1458, aa:
OTHER INFORMATION: (OTR)
US-07-862-588B-5

Query Match 52.0%; Score 13; DB 2; Length 1775;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 caggtgcacgcc 14
|||||
Db 934 CAGGTGCATGCC 946

RESULT 25
US-08-584-226-1/c
Sequence 1, Application US/08584226
Patent No. 5798240
GENERAL INFORMATION:
APPLICANT: Martinus, Susan A.
APPLICANT: Sassanitar, Mandana
APPLICANT: Kim, Sunghoon
APPLICANT: Lee, Sang Ho
APPLICANT: Schimmel, Paul R.
TITLE OF INVENTION: RECOMBINANT MYCOBACTERIAL METHIONYL-TRNA
TITLE OF INVENTION: SYNTHETASE GENES, TESTER STRAINS AND ASSAYS
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Militia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: USA
ZIP: 02173-4799
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/584,226
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/305,766
FILING DATE: 13-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Brook, David E.
REGISTRATION NUMBER: 22,592
REFERENCE/DOCKET NUMBER: CP194-052
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-6240

TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2290 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1563
US-08-584-226-1

Query Match 52.0%; Score 13; DB 1; Length 2290;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 tcgcagcagc 14
Db 1233 caggtgcattcc 1221

RESULT 26
US-09-134-566-15
; Sequence 15, Application US/09134566
; Patent No. 5998147
; GENERAL INFORMATION:
; APPLICANT: Petit, Christine
; APPLICANT: Denoyelle-Gryson, Francoise
; APPLICANT: Weil, Dominique
; APPLICANT: Marlin-Duvernois, Sandrine
; APPLICANT: Guesdon, Jean-Luc
; TITLE OF INVENTION: MUTATED POLYNUCLEOTIDE CORRESPONDING TO A MUTATION
; TITLE OF INVENTION: RESPONSIBLE FOR PRELINGUAL NON-SYNDROMIC DEAFNESS IN
; TITLE OF INVENTION: THE CONNEXIN 26 GENE AND METHOD OF DETECTING THIS
; FILE REFERENCE: 03495-0175
; CURRENT APPLICATION NUMBER: US/09/134,566
; CURRENT FILING DATE: 1998-08-14
; EARLIER APPLICATION NUMBER: 60/055,863
; EARLIER FILING DATE: 1998-08-15
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 15
; LENGTH: 2312
; TYPE: DNA
; ORGANISM: human connexin
US-09-134-566-15

Query Match 52.0%; Score 13; DB 2; Length 2312;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 tcgcagcagc 22
Db 960 tcgcagcagc 972

RESULT 27
US-09-134-566-8
; Sequence 8, Application US/09134566
; Patent No. 5998147
; GENERAL INFORMATION:
; APPLICANT: Petit, Christine
; APPLICANT: Denoyelle-Gryson, Francoise
; APPLICANT: Weil, Dominique
; APPLICANT: Marlin-Duvernois, Sandrine
; APPLICANT: Guesdon, Jean-Luc
; TITLE OF INVENTION: MUTATED POLYNUCLEOTIDE CORRESPONDING TO A MUTATION
; TITLE OF INVENTION: RESPONSIBLE FOR PRELINGUAL NON-SYNDROMIC DEAFNESS IN
; TITLE OF INVENTION: THE CONNEXIN 26 GENE AND METHOD OF DETECTING THIS

; TITLE OF INVENTION: HEREDITARY DEFECT
; FILE REFERENCE: 03495-0175
; CURRENT APPLICATION NUMBER: US/09/134,566
; CURRENT FILING DATE: 1998-08-14
; EARLIER APPLICATION NUMBER: 60/055,863
; EARLIER FILING DATE: 1998-08-15
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 8
; LENGTH: 2314
; TYPE: DNA
; ORGANISM: human connexin
US-09-134-566-8

Query Match 52.0%; Score 13; DB 2; Length 2314;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 tcgcagcagc 22
Db 962 tcgcagcagc 974

RESULT 28
US-08-759-444-2/c
; Sequence 2, Application US/08759444
; Patent No. 5824309
; GENERAL INFORMATION:
; APPLICANT: Dassarma, Shilditya
; APPLICANT: Moshed, Fazeela
; APPLICANT: Stuart, Elizabeth
; APPLICANT: Black, Samuel
; TITLE OF INVENTION: RECOMBINANT GAS VESICLES AND USES THEREOF
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/759,444
; FILING DATE: 05-DEC-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/008,200
; FILING DATE: 05-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Melkijohn, Ph.D., Anita L
; REGISTRATION NUMBER: 35,283
; REFERENCE/DOCKET NUMBER: 07880/003001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-542-5070
; TELEFAX: 617-542-8906
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8878 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
US-08-759-444-2

Query Match 52.0%; Score 13; DB 1; Length 8878;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 gtgcacgcagc 17
 |||
 Db 7684 GTGCATGCCAGC 7672

RESULT 29
 US-08-680-897-1/c
 ; Sequence 1, Application US/08680897
 ; Patent No. 6008051
 ; GENERAL INFORMATION:
 ; APPLICANT: Dassarma, Shiladitya
 ; APPLICANT: Ng, Mai-lap
 ; TITLE OF INVENTION: RECOMBINANT VECTOR AND PROCESS FOR CELL
 ; TITLE OF INVENTION: FLOTTATION
 ; NUMBER OF SEQUENCES: 1
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Scully, Scott, Murphy & Presser
 ; STREET: 400 Garden City Plaza
 ; CITY: Garden City
 ; STATE: New York
 ; COUNTRY: U.S.A.
 ; ZIP: 11530
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/680,897
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/271,270
 ; FILING DATE:
 ; APPLICATION NUMBER: 944,581
 ; FILING DATE: 14-SEP-1992
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Digilio, Frank S.
 ; REGISTRATION NUMBER: 31,346
 ; REFERENCE/DOCKET NUMBER: 8680
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (516) 742-4343
 ; TELEFAX: (516) 742-4366
 ; TELEX: 230 901 SANS UR
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 9880 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; US-08-680-897-1

Query Match 52.0%; Score 13; DB 3; Length 9880;
 Best Local Similarity 100.0%; Pred. No. 42;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 gtgcacgcagc 17
 |||
 Db 8684 GTGCATGCCAGC 8672

RESULT 30
 US-08-851-843A-92/c
 ; Sequence 92, Application US/08851843A
 ; Patent No. 6093809
 ; GENERAL INFORMATION:
 ; APPLICANT: Cecch, Thomas R.
 ; APPLICANT: Lingner, Joachim
 ; APPLICANT: Nakamura, Toru

APPLICANT: Chapman, Karen B.
 APPLICANT: Morin, Gregg B.
 APPLICANT: Harley, Calvin H.
 APPLICANT: Andrews, William H.
 TITLE OF INVENTION: NO. 6093809el Telomerase
 NUMBER OF SEQUENCES: 225
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Townsend and Townsend and Crew LLP
 STREET: Two Embarcadero Center, 8th Floor
 CITY: San Francisco
 STATE: California
 COUNTRY: United States of America
 ZIP: 94111
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/851,843A
 FILING DATE: 06-MAY-1997
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/846,017
 FILING DATE: 25-APR-1997
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/844,419
 FILING DATE: 18-APR-1997
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/724,643
 FILING DATE: 01-OCT-1996
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: Apple, Randolph T.
 REGISTRATION NUMBER: 36,429
 REFERENCE/DOCKET NUMBER: 015389-002930US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 576-0300
 TELEFAX: (415) 576-0300
 INFORMATION FOR SEQ ID NO: 92:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 19 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA
 US-08-851-843A-92

Query Match 48.0%; Score 12; DB 3; Length 19;
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 catgcgcagcat 19
 |||
 Db 19 CATGCCAGCAT 8

RESULT 31
 US-08-974-549A-384/c
 ; Sequence 384, Application US/08974549A
 ; Patent No. 6166178
 ; GENERAL INFORMATION:
 ; APPLICANT: Cecch, Thomas R.
 ; APPLICANT: Lingner, Joachim
 ; APPLICANT: Nakamura, Toru
 ; APPLICANT: Chapman, Karen B.
 ; APPLICANT: Morin, Gregg B.
 ; APPLICANT: Harley, Calvin H.
 ; APPLICANT: Andrews, William H.
 ; TITLE OF INVENTION: Human Telomerase Catalytic Subunit

NUMBER OF SEQUENCES: 727
 CORRESPONDENCE ADDRESS:
 ADDRESS: Townsend and Townsend and Crew LLP
 STREET: Two Embarcadero Center, Eighth Floor
 CITY: San Francisco
 STATE: California
 COUNTRY: USA
 ZIP: 94111-3834
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/974,549A
 FILING DATE: 19-NOV-1997
 CLASSIFICATION: 536
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/724,643
 FILING DATE: 01-OCT-1996
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/844,419
 FILING DATE: 18-APR-1997
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/846,017
 FILING DATE: 25-APR-1997
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/851,843
 FILING DATE: 06-MAY-1997
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/854,050
 FILING DATE: 09-MAY-1997
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/911,312
 FILING DATE: 14-AUG-1997
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/912,951
 FILING DATE: 14-AUG-1997
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/915,503
 FILING DATE: 14-AUG-1997
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: WO PCT/US97/17618
 FILING DATE: 01-OCT-1997
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: WO PCT/US97/17885
 FILING DATE: 01-OCT-1997
 ATTORNEY/AGENT INFORMATION:
 NAME: Apple, Randolph Ted
 REGISTRATION NUMBER: 36,429
 REFERENCE/DOCKET NUMBER: 015389-002610US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 576-0200
 TELEFAX: (415) 576-0300
 INFORMATION FOR SEQ ID NO: 384:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 19 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA
 FEATURE:
 NAME/KEY:
 LOCATION: 1..19
 OTHER INFORMATION: /note="TCPI.6 primer"
 US-08-974-549A-384

Query Match 48.0%; Score 12; DB 4; Length 19;
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 8 catgccagcat, 19

Db 19 CATGCCAGCAT 8
 |||||

RESULT 32
 US-08-373-124A-1692/C
 Sequence 1692, Application US/08373124A
 Patent No. 5646042
 GENERAL INFORMATION:
 APPLICANT: Stinchcomb, Dan T.
 APPLICANT: Draper, Kenneth
 APPLICANT: McSwigen, James
 APPLICANT: Jarvis, Thale
 TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR
 TITLE OF INVENTION: TREATMENT OF RESTENOSIS AND
 NUMBER OF SEQUENCES: 2627
 CORRESPONDENCE ADDRESS:
 ADDRESS: Lyon & Lyon
 STREET: 633 West Fifth Street
 CITY: Los Angeles
 STATE: California
 COUNTRY: U.S.A.
 ZIP: 90071
 COMPUTER READABLE FORM:
 MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
 MEDIUM TYPE: Storage
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: IBM P.C. DOS 5.0
 SOFTWARE: Word Perfect 5.1
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/373,124A
 FILING DATE: January 13, 1995
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/245,466
 FILING DATE: May 18, 1994
 APPLICATION NUMBER: 08/192,943
 FILING DATE: February 7, 1994
 APPLICATION NUMBER: 07/987,132
 FILING DATE: December 7, 1992
 APPLICATION NUMBER: 07/936,422
 FILING DATE: August 26, 1992
 ATTORNEY/AGENT INFORMATION:
 NAME: Wardburg, Richard
 REGISTRATION NUMBER: 32,327
 REFERENCE/DOCKET NUMBER: 209/035
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (213) 489-1600
 TELEFAX: (213) 955-0440
 TELEX: 67-3510
 INFORMATION FOR SEQ ID NO: 1692:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 38 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 US-08-373-124A-1692

Query Match 48.0%; Score 12; DB 1; Length 38;
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 gccagcatcg 23
 |||||

Db 38 GCCAGCATTCG 27
 |||||

RESULT 33
 US-08-435-628-1692/C
 Sequence 1692, Application US/08435628
 Patent No. 5817796

GENERAL INFORMATION:
APPLICANT: Stinchcomb, Dan T.
APPLICANT: Draper, Kenneth
APPLICANT: McSwigen, James
APPLICANT: Jarvis, Thale
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR
TITLE OF INVENTION: TREATMENT OF RESTENOSIS AND
TITLE OF INVENTION: CANCER USING RIBOZYMES
NUMBER OF SEQUENCES: 2627
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 MB
MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/435,628
FILING DATE: 05-MAY-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/373,124
FILING DATE: January 13, 1995
APPLICATION NUMBER: 08/245,466
FILING DATE: May 18, 1994
APPLICATION NUMBER: 08/192,943
FILING DATE: February 7, 1994
APPLICATION NUMBER: 07/987,132
FILING DATE: December 7, 1992
APPLICATION NUMBER: 07/936,422
FILING DATE: August 26, 1992
ATTORNEY/AGENT INFORMATION:
NAME: Wardburg, Richard
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 209/035
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 1692:
SEQUENCE CHARACTERISTICS:
LENGTH: 38 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-435-628-1692

Query Match 48.0%; Score 12; DB 1; Length 38;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 gccagatttcg 23
|||||
DB 38 gccagatttcg 27

RESULT 34
US-09-023-228B-59
Sequence 59, Application US/090232228B
Patent No. 6140490
GENERAL INFORMATION:
APPLICANT: BIESSECKER, GREGORY
APPLICANT: GOLD, LARRY
TITLE OF INVENTION: HIGH AFFINITY NUCLEIC ACID LIGANDS OF
TITLE OF INVENTION: COMPLEMENT SYSTEM PROTEINS

NUMBER OF SEQUENCES: 157
CORRESPONDENCE ADDRESS:
ADDRESSEE: Swanson & Bratschun, L.L.C.
STREET: 8400 E. Prentice Place #200
CITY: Denver
STATE: Colorado
COUNTRY: US
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Wordperfect 8.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/023,228B
FILING DATE: 12-FEBRUARY-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US97/01739
FILING DATE: 30 JAN 1997
APPLICATION NUMBER: 08/595,335
FILING DATE: 1 FEB 1996
ATTORNEY/AGENT INFORMATION:
NAME: Barry J. Swanson, Esq.
REGISTRATION NUMBER: 33,215
REFERENCE/DOCKET NUMBER: NEX50/CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 793-3333
TELEFAX: (303) 793-3433
INFORMATION FOR SEQ ID NO: 59:
SEQUENCE CHARACTERISTICS:
LENGTH: 61 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: RNA
FEATURE:
OTHER INFORMATION: All pyrimidines are 2'-F modified
US-09-023-228B-59

Query Match 48.0%; Score 12; DB 3; Length 61;
Best Local Similarity 66.7%; Pred. No. 1.4e+02;
Matches 8; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 14 gccatttcgat 25
|||||
DB 23 CAGCAUUCGAU 34

RESULT 35
US-08-488-402A-16/C
Sequence 16, Application US/08488402A
Patent No. 5837456
GENERAL INFORMATION:
APPLICANT: GOLD ET AL.
TITLE OF INVENTION: HIGH AFFINITY OLIGONUCLEOTIDE LIGANDS TO
TITLE OF INVENTION: CHORIONIC GONADOTROPIN HORMONE AND RELATED GLYCOPROTEIN
TITLE OF INVENTION: HORMONES
NUMBER OF SEQUENCES: 160
CORRESPONDENCE ADDRESS:
ADDRESSEE: Swanson & Bratschun, L.L.C.
STREET: 8400 E. Prentice Avenue, Suite 200
CITY: Englewood
STATE: Colorado
COUNTRY: USA
ZIP: 80111
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3 1/2 diskette, 1.44 MG
COMPUTER: IBM PC compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/488,402A
FILING DATE: 07 JUNE 1995
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/536,428
FILING DATE: 11-JUNE-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/714,131
FILING DATE: 10-JUNE-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/931,473
FILING DATE: 17-AUGUST-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/964,624
FILING DATE: 21-OCTOBER-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/177,991
FILING DATE: 8-SEPTEMBER-1993
ATTORNEY/AGENT INFORMATION:
NAME: Barry J. Swanson
REGISTRATION NUMBER: 33,215
REFERENCE/DOCKET NUMBER: NEX36-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 793-3333
TELEFAX: (303) 793-3433
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 69 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
OTHER INFORMATION: All C's are 2'-NH2 modified
FEATURE:
OTHER INFORMATION: All U's are 2'-NH2 modified
US-08-488-402A-16

Query Match 48.0%; Score 12; DB 2; Length 69;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 aggtgcacgcc 14
|||||
Db 25 AGTGCATCGCC 14

RESULT 36
US-08-484-552A-16/C
Sequence 16, Application US/08484552A
Patent No. 5849890
GENERAL INFORMATION:
APPLICANT: GOLD, LARRY
APPLICANT: JAYASENA, SUMEDHA
APPLICANT: NIEUWLANDT, DAN
APPLICANT: DAVIS, KEN
TITLE OF INVENTION: HIGH AFFINITY OLIGONUCLEOTIDE LIGANDS TO CHORIONIC
TITLE OF INVENTION: GONADOTROPIN HORMONE AND RELATED GLYCOPROTEIN HORMONES
NUMBER OF SEQUENCES: 160
CORRESPONDENCE ADDRESS:
ADDRESSEE: Swanson & Bratschun, L.L.C.
STREET: 8400 E. Prentice Avenue, Suite 200
CITY: Englewood
STATE: Colorado
COUNTRY: USA
ZIP: 80111
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3 1/2 diskette, 1.44 MG
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,552A

FILING DATE: 07 JUNE 1995
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/536,428
FILING DATE: 11-JUNE-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/714,131
FILING DATE: 10-JUNE-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/931,473
FILING DATE: 17-AUGUST-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/964,624
FILING DATE: 21-OCTOBER-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/177,991
FILING DATE: 8-SEPTEMBER-1993
ATTORNEY/AGENT INFORMATION:
NAME: Barry J. Swanson
REGISTRATION NUMBER: 33,215
REFERENCE/DOCKET NUMBER: NEX36-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 793-3333
TELEFAX: (303) 793-3433
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 69 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
OTHER INFORMATION: All C's are 2'-NH2 modified
FEATURE:
OTHER INFORMATION: All U's are 2'-NH2 modified
US-08-484-552A-16

Query Match 48.0%; Score 12; DB 2; Length 69;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 aggtgcacgcc 14
|||||
Db 25 AGTGCATCGCC 14

RESULT 37
PCT-US96-09472-16/C
Sequence 16, Application PC/TUS9609472
GENERAL INFORMATION:
APPLICANT: GOLD ET AL.
TITLE OF INVENTION: HIGH AFFINITY OLIGONUCLEOTIDE
TITLE OF INVENTION: LIGANDS TO CHORIONIC GONADOTROPIN HORMONE AND RELATED
TITLE OF INVENTION: GLYCOPROTEIN HORMONES
NUMBER OF SEQUENCES: 160
CORRESPONDENCE ADDRESS:
ADDRESSEE: Swanson & Bratschun, L.L.C.
STREET: 8400 E. Prentice Avenue, Suite 200
CITY: Englewood
STATE: Colorado
COUNTRY: USA
ZIP: 80111
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3 1/2 diskette, 1.44 MG
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/09472
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA: 08/488,402
APPLICATION NUMBER:

FILED DATE: 07 JUNE 1995
PRIOR APPLICATION DATA: 08/484,552
APPLICATION NUMBER: 08/484,552
FILING DATE: 07 JUNE 1995
ATTORNEY/AGENT INFORMATION:
NAME: Barry J. Swanson
REGISTRATION NUMBER: 33,215
REFERENCE/DOCKET NUMBER: NEX36-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 793-3333
TELEFAX: (303) 793-3433
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 69 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
OTHER INFORMATION: All C's are 2'-NH2 modified
FEATURE:
OTHER INFORMATION: All U's are 2'-NH2 modified
PCT-US96-09472-16

Query Match 48.0%; Score 12; DB 5; Length 69;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 3 aggtcgcgcgc 14
|||||
Db 25 AGTGCATCGCC 14

RESULT 38
US-08-998-416-93/c
Sequence 93, Application US/0898416
Patent No. 6239264
GENERAL INFORMATION:
APPLICANT: Philippsen, Peter
APPLICANT: Pohlmann, Rainer
APPLICANT: Steiner, Sabine
APPLICANT: Mohr, Christine
APPLICANT: Wendland, Jurgen
APPLICANT: Knechtel, Philipp
APPLICANT: Redischung, Corinne
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPTII
TITLE OF INVENTION: AND USES THEREOF
NUMBER OF SEQUENCES: 1152
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6239264artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: No. 6239264th Carolina
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/998,416
FILING DATE: 24-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587

TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 93:
SEQUENCE CHARACTERISTICS:
LENGTH: 386 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: PAG1031UP
US-08-998-416-93

Query Match 48.0%; Score 12; DB 4; Length 386;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 6 tgcgcgcgcgc 17
|||||
Db 292 TGCATCGCCGCGC 281

RESULT 39
US-08-743-637B-184/c
Sequence 184, Application US/08743637B
Patent No. 5994066
GENERAL INFORMATION:
APPLICANT: BERGERON, Michel G.
APPLICANT: PICARD, Francois J.
APPLICANT: OUELLETTE, Marc
APPLICANT: ROY, Paul H.
TITLE OF INVENTION: SPECIES-SPECIFIC AND UNIVERSAL DNA
TITLE OF INVENTION: PROBES AND AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND
IDENTIFY COMMON BACTERIAL PATHOGENS AND ASSOCIATED
TITLE OF INVENTION: ANTIBIOTIC RESISTANCE GENES FROM CLINICAL SPECIMENS ...
NUMBER OF SEQUENCES: 273
CORRESPONDENCE ADDRESS:
ADDRESSEE: OUBARLES & BRADY
STREET: 411 EAST WISCONSIN AVENUE
CITY: MILWAUKEE
STATE: WISCONSIN
COUNTRY: USA
ZIP: 53202-4497
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/743,637B
FILING DATE: 04-NOV-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/526,840
FILING DATE: 11-SEP-1995
ATTORNEY/AGENT INFORMATION:
NAME: BAKER, Jean C.
REGISTRATION NUMBER: 35,433
REFERENCE/DOCKET NUMBER: 850586.90012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (414) 277-5591
TELEFAX: (414) 277-5000
INFORMATION FOR SEQ ID NO: 184:
SEQUENCE CHARACTERISTICS:
LENGTH: 438 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
US-08-743-637B-184

Query Match 48.0%; Score 12; DB 2; Length 438;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcaggtgcatcg 12
|||||
DB 28 GCAGGTGCATCG 17

RESULT 40

US-08-998-416-56/c
Sequence 56, Application US/08998416
Patent No. 6239264

GENERAL INFORMATION:

APPLICANT: Philippsen, Peter
APPLICANT: Pohlmann, Rainer

APPLICANT: Steiner, Sabine
APPLICANT: Mohr, Christine

APPLICANT: Wendland, Jurgen
APPLICANT: Knechtle, Philipp

APPLICANT: Reibischung, Corinne
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSTYPII

TITLE OF INVENTION: AND USES THEREOF
NUMBER OF SEQUENCES: 1152

CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6239264artis Corporation

STREET: 3054 Cornwallis Road
CITY: Research Triangle Park

STATE: No. 6239264th Carolina
COUNTRY: USA

ZIP: 27709

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/998,416
FILING DATE: 24-DEC-1997

CLASSIFICATION: 435
PRIOR APPLICATION DATA:

APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996

ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy

REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976

TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587

TELEFAX: 919-541-8589
INFORMATION FOR SEQ ID NO: 56:

SEQUENCE CHARACTERISTICS:
LENGTH: 656 base pairs

TYPE: nucleic acid
STRANDEDNESS: single

TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)

ORIGINAL SOURCE:
ORGANISM: PAG1016RP

US-08-998-416-56

Query Match 48.0%; Score 12; DB 4; Length 656;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 tgcattgcacgc 17
|||||
DB 53 TGCATTGCCACGC 42

RESULT 41

US-08-526-840B-175/c
Sequence 175, Application US/08526840B
Patent No. 6001564

GENERAL INFORMATION:
APPLICANT: BERGERON, Michel G.

APPLICANT: PICARD, Francois J.
APPLICANT: OUELLETTE, Marc

APPLICANT: Roy, Paul H.
TITLE OF INVENTION: SPECIES-SPECIFIC AND UNIVERSAL DNA

TITLE OF INVENTION: PROBES AND AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND
TITLE OF INVENTION: IDENTIFY COMMON BACTERIAL PATHOGENS AND ASSOCIATED

TITLE OF INVENTION: ANTIBIOTIC RESISTANCE GENES FROM CLINICAL SPECIMENS ...
NUMBER OF SEQUENCES: 273

CORRESPONDENCE ADDRESS:
ADDRESSEE: QUARLES & BRADY

STREET: 411 EAST WISCONSIN AVENUE
CITY: MILWAUKEE

STATE: WISCONSIN
COUNTRY: USA

ZIP: 53202-4497

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/743,637B
FILING DATE: 04-NOV-1996

CLASSIFICATION: 435
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 35,433
REFERENCE/DOCKET NUMBER: 850586, 90012

TELECOMMUNICATION INFORMATION:
TELEPHONE: (414) 277-5000

TELEFAX: (414) 277-5591
INFORMATION FOR SEQ ID NO: 175:

SEQUENCE CHARACTERISTICS:
LENGTH: 660 base pairs

TYPE: nucleic acid
STRANDEDNESS: double

TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)

US-08-743-637B-175

Query Match 48.0%; Score 12; DB 2; Length 660;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 cagatttcgat 25
|||||
DB 627 CAGATTTCGAT 616

RESULT 42

US-08-526-840B-175/c
Sequence 175, Application US/08526840B
Patent No. 6001564

GENERAL INFORMATION:
APPLICANT: BERGERON, Michel G.

APPLICANT: PICARD, Francois J.
APPLICANT: OUELLETTE, Marc

APPLICANT: Roy, Paul H.
TITLE OF INVENTION: SPECIES-SPECIFIC AND UNIVERSAL PROBES AND

TITLE OF INVENTION: AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND IDENTIFY
TITLE OF INVENTION: COMMON BACTERIAL PATHOGENS AND ANTIBIOTIC RESISTANCE GENES

TITLE OF INVENTION: FROM CLINICAL SPECIMENS FOR ROUTINE DIAGNOSIS IN ...
NUMBER OF SEQUENCES: 177

CORRESPONDENCE ADDRESS:
ADDRESSEE: QUARLES & BRADY

STREET: 411 EAST WISCONSIN AVENUE
CITY: MILWAUKEE

STATE: WISCONSIN
COUNTRY: USA

ZIP: 53202-4497

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/743,637B
FILING DATE: 04-NOV-1996

CLASSIFICATION: 435
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 35,433
REFERENCE/DOCKET NUMBER: 850586, 90012

TELECOMMUNICATION INFORMATION:
TELEPHONE: (414) 277-5000

TELEFAX: (414) 277-5591
INFORMATION FOR SEQ ID NO: 175:

SEQUENCE CHARACTERISTICS:
LENGTH: 660 base pairs

TYPE: nucleic acid
STRANDEDNESS: double

TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)

US-08-743-637B-175

STREET: 411 East Wisconsin Avenue
CITY: Milwaukee
STATE: Wisconsin
COUNTRY: USA
ZIP: 53202-4497
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/526,840B
FILING DATE: 11-SEP-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/304,732
FILING DATE: 12-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: BAKER, Jean C.
REGISTRATION NUMBER: 35,433
REFERENCE/DOCKET NUMBER: 850586,90012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (414) 277-5000
TELEFAX: (414) 277-5591
INFORMATION FOR SEQ ID NO: 175:
SEQUENCE CHARACTERISTICS:
LENGTH: 660 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-526-840B-175

Query Match 48.0%; Score 12; DB 3; Length 660;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 14 cagcattcgat 25
|||||
DB 627 CAGCATTTCGAT 616

RESULT 43
US-08-998-416-767
Sequence 767, Application US/08998416
Patent No. 6239264
GENERAL INFORMATION:
APPLICANT: Philippsen, Peter
APPLICANT: Pohlmann, Rainer
APPLICANT: Steiner, Sabine
APPLICANT: Mohr, Christine
APPLICANT: Wendland, Jurgen
APPLICANT: Knechtle, Philipp
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPHII
TITLE OF INVENTION: AND USES THEREOF
NUMBER OF SEQUENCES: 1152
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6239264artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: No. 6239264th Carolina
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/998,416
FILING DATE: 24-DEC-1997

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 767:
SEQUENCE CHARACTERISTICS:
LENGTH: 715 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: PAG1495RP
US-08-998-416-767

Query Match 48.0%; Score 12; DB 4; Length 715;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 10 tcgccagcattc 21
|||||
DB 356 TCGCCAGCATTT 367

RESULT 44
US-08-998-416-237/C
Sequence 237, Application US/08998416
Patent No. 6239264
GENERAL INFORMATION:
APPLICANT: Philippsen, Peter
APPLICANT: Pohlmann, Rainer
APPLICANT: Steiner, Sabine
APPLICANT: Mohr, Christine
APPLICANT: Wendland, Jurgen
APPLICANT: Knechtle, Philipp
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPHII
TITLE OF INVENTION: AND USES THEREOF
NUMBER OF SEQUENCES: 1152
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6239264artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: No. 6239264th Carolina
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/998,416
FILING DATE: 24-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689

;/ INFORMATION FOR SEQ ID NO: 237:
;/ SEQUENCE CHARACTERISTICS:
;/ LENGTH: 754 base pairs
;/ TYPE: nucleic acid
;/ STRANDEDNESS: single
;/ TOPOLOGY: linear
;/ MOLECULE TYPE: DNA (genomic)
;/ ORIGINAL SOURCE:
;/ ORGANISM: PAG1215RP
;/ US-08-998-416-237

Query Match 48.0%; Score 12; DB 4; Length 754;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 caggtcgcacgc 13
|||||
Db 461 CAGGTGCATCGC 450

RESULT 45
US-08-990-823-39
; Sequence 39, Application US/08990823D
; Patent No. 6228371
; GENERAL INFORMATION:
; APPLICANT: Nano, Francis
; TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
; FILE REFERENCE: 49086
; CURRENT APPLICATION NUMBER: US/08/990,823D
; EARLIER APPLICATION NUMBER: US 96/10375
; EARLIER FILING DATE: 1996-06-14
; EARLIER APPLICATION NUMBER: 60/000,254
; EARLIER FILING DATE: 1995-06-15
; NUMBER OF SEQ ID NOS: 113
; SOFTWARE: Patentlin Ver. 2.0
; SEQ ID NO 39
; LENGTH: 841
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; NAME/KEY: Modified base
; OTHER INFORMATION: n represents a or g or c or t/u
US-08-990-823-39

Query Match 48.0%; Score 12; DB 4; Length 841;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 gcacgcacgca 18
|||||
Db 370 gcacgcacgca 381

Search completed: October 9, 2001, 15:55:36
Job time: 13212 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 15:15:31 ; Search time 5323.87 Seconds
(without alignments)
44.389 Million cell updates/sec

Title: US-09-396-196f-6

Perfect score: 25
Sequence: 1 gaagtgatcgccagcatcgcgat 25

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 10228115 seqs, 4726426750 residues

Word size : 9

Total number of hits satisfying chosen parameters: 443280

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

EST:*

1: gb_est1.*
2: gb_est2.*
3: gb_est3.*
4: gb_est4.*
5: gb_est5.*
6: gb_est6.*
7: gb_est7.*
8: gb_est8.*
9: gb_est9.*
10: gb_est10.*
11: gb_est11.*
12: gb_est12.*
13: gb_est13.*
14: gb_est14.*
15: gb_est15.*
16: gb_est16.*
17: gb_est17.*
18: gb_est18.*
19: gb_est19.*
20: gb_est20.*
21: gb_est21.*
22: gb_est22.*
23: gb_est23.*
24: gb_est24.*
25: gb_est25.*
26: gb_est26.*
27: gb_est27.*
28: gb_est28.*
29: gb_est29.*
30: gb_est30.*
31: gb_est31.*
32: gb_est32.*
33: gb_est33.*
34: gb_est34.*
35: gb_est35.*
36: gb_est36.*
37: gb_est37.*
38: gb_est38.*
39: gb_est39.*
40: gb_est40.*
41: gb_est41.*
42: gb_est42.*
43: gb_est43.*
44: gb_est44.*
45: gb_est45.*
46: gb_est46.*
47: gb_est47.*
48: gb_est48.*
49: gb_est49.*
50: gb_est50.*
51: gb_est51.*
52: gb_est52.*
53: gb_est53.*
54: gb_est54.*
55: gb_est55.*
56: gb_est56.*
57: gb_est57.*
58: gb_est58.*
59: gb_est59.*
60: gb_est60.*
61: gb_est61.*
62: gb_est62.*
63: gb_est63.*
64: gb_est64.*
65: gb_est65.*
66: gb_est66.*
67: gb_est67.*
68: gb_est68.*
69: gb_est69.*
70: gb_est70.*
71: gb_est71.*
72: gb_est72.*
73: gb_est73.*
74: gb_est74.*
75: gb_est75.*
76: gb_est76.*
77: gb_est77.*
78: gb_est78.*
79: gb_est79.*
80: gb_est80.*
81: gb_est81.*
82: gb_est82.*
83: gb_est83.*
84: gb_est84.*
85: gb_est85.*
86: gb_est86.*
87: gb_est87.*
88: gb_est88.*
89: gb_est89.*
90: gb_est90.*
91: gb_est91.*
92: gb_est92.*
93: gb_est93.*
94: gb_est94.*
95: gb_est95.*
96: gb_est96.*
97: gb_est97.*
98: gb_est98.*
99: gb_est99.*
100: gb_est100.*
101: gb_est101.*
102: gb_est102.*
103: gb_est103.*
104: gb_est104.*
105: gb_est105.*
106: gb_est106.*
107: gb_est107.*
108: gb_est108.*
109: gb_est109.*
110: gb_est110.*
111: gb_est111.*
112: gb_est112.*
113: gb_est113.*
114: gb_est114.*
115: gb_est115.*
116: gb_est116.*

44: em_esthum10.*
45: em_esthum11.*
46: em_esthum12.*
47: em_esthum13.*
48: em_esthum14.*
49: em_esthum15.*
50: em_esthum16.*
51: em_esthum17.*
52: em_esthum18.*
53: em_esthum19.*
54: em_esthum20.*
55: em_esthum21.*
56: em_esthum22.*
57: em_esthum23.*
58: em_esthum24.*
59: em_esthum25.*
60: em_esthum26.*
61: em_esthum27.*
62: em_esthum28.*
63: em_esthum29.*
64: em_esthum30.*
65: em_esthum31.*
66: em_esthum32.*
67: em_esthum33.*
68: em_esthum34.*
69: em_esthum35.*
70: em_esthum36.*
71: em_esthum37.*
72: em_esthum38.*
73: em_esthum39.*
74: em_esthum40.*
75: em_esthum41.*
76: em_esthum42.*
77: em_esthum43.*
78: em_esthum44.*
79: em_esthum45.*
80: em_esthum46.*
81: em_esthum47.*
82: em_esthum48.*
83: em_esthum49.*
84: em_esthum50.*
85: em_esthum51.*
86: em_esthum52.*
87: em_esthum53.*
88: em_esthum54.*
89: em_esthum55.*
90: em_esthum56.*
91: em_esthum57.*
92: em_esthum58.*
93: em_esthum59.*
94: em_esthum60.*
95: em_esthum61.*
96: em_esthum62.*
97: em_esthum63.*
98: em_esthum64.*
99: em_esthum65.*
100: em_esthum66.*
101: em_esthum67.*
102: em_esthum68.*
103: em_esthum69.*
104: em_esthum70.*
105: em_esthum71.*
106: em_esthum72.*
107: em_esthum73.*
108: em_esthum74.*
109: em_esthum75.*
110: em_esthum76.*
111: em_esthum77.*
112: em_esthum78.*
113: em_esthum79.*
114: em_esthum80.*
115: em_esthum81.*
116: em_esthum82.*

```
117: gb_est148:*
118: gb_est149:*
119: gb_est150:*
120: gb_est151:*
121: gb_est152:*
122: gb_est153:*
123: gb_est154:*
124: gb_est155:*
125: gb_est156:*
126: gb_est157:*
127: gb_est158:*
128: gb_est159:*
129: gb_est160:*
130: gb_est161:*
131: gb_est162:*
132: gb_est163:*
133: gb_est164:*
134: gb_est165:*
135: gb_est166:*
136: gb_est167:*
137: gb_est168:*
138: gb_est169:*
139: gb_est170:*
140: gb_est171:*
141: gb_est172:*
142: gb_est173:*
143: gb_est174:*
144: gb_est175:*
145: gb_est176:*
146: gb_est177:*
147: gb_est178:*
148: gb_est179:*
149: gb_est180:*
150: gb_est181:*
151: gb_est182:*
152: gb_est183:*
153: gb_est184:*
154: gb_est185:*
155: gb_est186:*
156: gb_est187:*
157: gb_est188:*
158: gb_est189:*
159: gb_est190:*
160: gb_est191:*
161: gb_est192:*
162: gb_est193:*
163: gb_est194:*
164: gb_est195:*
165: gb_est196:*
166: gb_est197:*
167: gb_est198:*
168: gb_est199:*
169: gb_est200:*
170: gb_est201:*
171: gb_est202:*
172: gb_est203:*
173: gb_est204:*
174: gb_est205:*
175: gb_est206:*
176: gb_est207:*
177: gb_est208:*
178: gb_est209:*
179: gb_est210:*
180: gb_est211:*
181: gb_est212:*
182: gb_est213:*
183: gb_est214:*
184: gb_est215:*
185: gb_est216:*
186: gb_est217:*
187: gb_est218:*
188: gb_est219:*
189: gb_est220:*

190: gb_est110:*
191: gb_est111:*
192: gb_est112:*
193: gb_est113:*
194: gb_est114:*
195: gb_est115:*
196: gb_est116:*
197: gb_est117:*
198: gb_est118:*
199: gb_est119:*
200: gb_est120:*
201: gb_est121:*
202: gb_est122:*
203: gb_est123:*
204: gb_est124:*
205: gb_est125:*
206: gb_est126:*
207: gb_est127:*
208: gb_est128:*
209: gb_est129:*
210: gb_est130:*
211: gb_est131:*
212: gb_est132:*
213: gb_est133:*
214: gb_est134:*
215: gb_est135:*
216: gb_est136:*
217: gb_est137:*
218: gb_est138:*
219: gb_est139:*
220: gb_est140:*
221: gb_est141:*
222: gb_est142:*
223: gb_est143:*
224: gb_est144:*
225: gb_est145:*
226: gb_est146:*
227: gb_est147:*
228: gb_est148:*
229: gb_est149:*
230: gb_est150:*
231: gb_est151:*
232: gb_est152:*
233: gb_est153:*
234: gb_est154:*
235: gb_est155:*
236: gb_est156:*
237: gb_est157:*
238: gb_est158:*
239: gb_est159:*
240: gb_est160:*
241: gb_est161:*
242: gb_est162:*
243: gb_est163:*
244: gb_est164:*
245: gb_est165:*
246: gb_est166:*
247: gb_est167:*
248: gb_est168:*
249: gb_est169:*
250: gb_est170:*
251: gb_est171:*
252: gb_est172:*
253: gb_est173:*
254: gb_est174:*
255: gb_est175:*
256: gb_est176:*
257: gb_est177:*
258: gb_est178:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	17	68.0	498	21	AI532390 SD03891.5
C 2	16	68.0	376	171	BF933544 CMI-NT027
C 3	16	64.0	399	156	D24778 R1C2545A R
C 4	16	64.0	429	107	A0030894 A0030894
C 5	16	64.0	510	107	A0031475 A0031475
C 6	16	64.0	548	258	P968R A0390674 Leishman1
C 7	16	64.0	666	107	A0030895 A0030895
C 8	16	64.0	704	107	A0078750 A0078750
C 9	16	64.0	751	170	BF864751 963054E05
C 10	15	60.0	228	246	A578088 18f02_Sho
C 11	15	60.0	266	19	A1410504 EST338797
C 12	15	60.0	338	162	BE065518 RC3-BT031
C 13	15	60.0	341	31	AV551712 AV551712
C 14	15	60.0	380	12	AA800788 EST190285
C 15	15	60.0	393	158	HE0965 YF22C07.r1
C 16	15	60.0	394	189	T86053 Y662405.r1
C 17	15	60.0	396	167	BE429060 YMD013.H0
C 18	15	60.0	399	162	BE044431 hc45d04.x
C 19	15	60.0	414	18	A1268648 q039f05.x
C 20	15	60.0	435	223	A0027122 C1T-HSP-2
C 21	15	60.0	437	158	HE9200 YF95f10.r1
C 22	15	60.0	438	187	N92081 2a23h03.r1
C 23	15	60.0	445	32	AV722315 AV722315
C 24	15	60.0	453	108	A0166733 A0166733
C 25	15	60.0	463	167	BE446670 WHE1139_F
C 26	15	60.0	466	1	AA026650 2199e08.r
C 27	15	60.0	475	12	AA821080 GM09622.5
C 28	15	60.0	475	158	HE0979 YF22f01.r1
C 29	15	60.0	475	173	BE101817 RH1Z2_22
C 30	15	60.0	491	23	AI692369 wd63f02.x
C 31	15	60.0	491	146	BF211097 601812706
C 32	15	60.0	499	171	BF916376 IL3-UT011
C 33	15	60.0	505	156	C98143 C98143 Rice
C 34	15	60.0	507	189	T84279 Y447D04.r1
C 35	15	60.0	514	154	BE467085 1A04C06 B
C 36	15	60.0	524	104	AI957283 ul85e12.x
C 37	15	60.0	528	158	H74120 Y815f08.r1
C 38	15	60.0	536	174	BE232022 na635e06
C 39	15	60.0	539	170	BF839030 PM2-HT035
C 40	15	60.0	545	237	A2071920 RPCI-23-3
C 41	15	60.0	547	118	AMS88900 ra06905.Y
C 42	15	60.0	553	241	AZ283519 RPCI-23-1
C 43	15	60.0	567	156	C98142 C98142 Rice
C 44	15	60.0	581	7	AA405993 zu66f09.r
C 45	15	60.0	605	233	AO783596 HS_3121_A

ALIGNMENTS

RESULT	1	498 bp	EST	18-MAR-1999
AI532390/c				
LOCUS				
DEFINITION	SD03891.5prime SD Drosophila melanogaster Schneider L2 cell culture			
ACCESION	AI532390			
VERSION	AI532390.1	GI:4446525		
KEYWORDS	EST.			
SOURCE	fruit fly.			
ORGANISM	Drosophila melanogaster			
REFERENCE	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.			
AUTHORS	Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C., Brokstein			
	P., Lewis,S. and Rubin,G.M.			

TITLE BDGP/HIMI Drosophila EST Project
JOURNAL Unpublished (1997)
COMMENT Contact: Harvey, D.
G. M. Rubin-Molecular and Cell Biology
University of California Berkeley
539 LSA, Berkeley, CA 94720-3300, USA
Fax: 510 643 9947
Email: <http://www.fruitfly.org/EST/estefruitfly.berkeley.edu>
Plate: 38 row: H column: 7
High quality sequence stop: 433.

FEATURES
source
1..498
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="SD03891"
/lab_host="SD Drosophila melanogaster Schneider L2 cell culture port2"
/lab_host="DH5-alpha"
/note="Vector: port2; Site1: EcoRI; Site2: XhoI; sized fractionated cDNAs were directly ligated into port2. Plasmid cDNA library."
Plasmid cDNA library.

BASE COUNT 158 a 106 c 124 g 110 t
ORIGIN

Query Match 68.0%; Score 17; DB 21; Length 498;
Best local Similarity 100.0%; Pred. No. 4.7;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 7 gcacgcacgacatccg 23
|||||
Db 266 GCATGCCGACGATTTCG 250

RESULT 2
BF933544/c 376 bp mRNA EST 22-JAN-2001
LOCUS CMI-NT0270-211200-661-g03 NT0270 Homo sapiens CDNA, mRNA sequence.
DEFINITION BF933544
VERSION BF933544.1 GI:12350868
KEYWORDS EST.

SOURCE
ORGANISM Homo sapiens
human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE
Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 376)

AUTHORS
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bata,G.S., Simpson,D.H.,
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.

TITLE
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE
20202663
COMMENT
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil

FEATURES
source
1..376
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(<http://www.ludwig.org.br/scripts/gethtml2.pl?l1=CMI6t2-CMI-NT0270-211200-661-g03&t3=2000-12-21&t4=1>)
Seq primer: puc 18 forward
High quality sequence start: 10
High quality sequence stop: 376.
Location/Qualifiers
1..376

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NT0270"
/dev_stage="Adult"
/note="Organ: nervous_tumor; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT      77 a      104 c      99 g      96 t
ORIGIN

Query Match      64.0%; Score 16; DB 171; Length 376;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 ggtgcacgcacgcat 19
|||||
Db 329 GGTGATCGCCAGCAT 314

RESULT 3
D24778 399 bp mRNA EST 08-JUL-1999
DEFINITION R1CR2545A Rice root Oryza sativa cDNA clone R2545_1A, mRNA
sequence.
ACCESSION D24778
VERSION D24778.1 GI:428626
KEYWORDS EST.
SOURCE Oryza sativa.
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
1 (bases 1 to 399)
REFERENCE
AUTHORS Minobe,Y. and Sasaki,T.
TITLE Rice cDNA from root
JOURNAL Unpublished (1995)
CONTACT: Takuji Sasaki
NATIONAL INSTITUTE OF AGROBIOLOGICAL RESOURCES
RICE GENOME RESEARCH PROGRAM
2-1-2 Kannondai,Tsukuba
Ibaraki,
Japan 305
Tel: 0298-38-7441
Fax: 0298-38-7468
Email: tsasaki@abr.affrc.go.jp
PROJECT="RGP".

FEATURES
source location/Qualifiers
1..399
/organism="Oryza sativa"
/strain="Nipponbare, sub_species Japonica"
/db_xref="taxon:4530"
/clone="R2545_1A"
/clone_lib="Rice root"
/note="Prepared from seedling root."
BASE COUNT 114 a 70 c 102 g 112 t 1 others
ORIGIN

Query Match      64.0%; Score 16; DB 156; Length 399;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 ggtgcacgcacgcat 19
|||||
Db 251 GGTGATCGCCAGCAT 266

RESULT 4

```

```

AU030894/c
LOCUS AU030894 429 bp mRNA EST 29-OCT-1998
DEFINITION AU030894 Rice cDNA from immature leaf including apical meristem
Oryza sativa cDNA clone E60405_1A, mRNA sequence.
ACCESSION AU030894
VERSION AU030894.1 GI:376784
KEYWORDS EST.
SOURCE Oryza sativa.
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
1 (bases 1 to 429)
REFERENCE
AUTHORS Sasaki,T. and Yamamoto,K.
TITLE Rice cDNA from immature leaf including apical meristem
JOURNAL Unpublished (1997)
CONTACT: Takuji Sasaki
NATIONAL INSTITUTE OF AGROBIOLOGICAL RESOURCES
RICE GENOME RESEARCH PROGRAM
2-1-2 Kannondai,Tsukuba
Ibaraki,
Japan 305
Tel: 0298-38-7441
Fax: 0298-38-7468
Email: tsasaki@abr.affrc.go.jp
PROJECT="RGP".

FEATURES
source location/Qualifiers
1..429
/organism="Oryza sativa"
/strain="Nipponbare"
/db_xref="taxon:4530"
/clone="E60405_1A"
/clone_lib="Rice cDNA from immature leaf including apical
meristem"
/dev_stage="Immature"
/note="Organ: leaf; Immature leaf including apical
meristem (under long day condition)."
BASE COUNT 106 a 122 c 101 g 99 t 1 others
ORIGIN

Query Match      64.0%; Score 16; DB 107; Length 429;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 catgcacgcatctcg 23
|||||
Db 302 CATGCCAGCATTCG 287

RESULT 5
AU031475/c
LOCUS AU031475 510 bp mRNA EST 30-OCT-1998
DEFINITION AU031475 Rice cDNA from immature leaf including apical meristem
Oryza sativa cDNA clone E61659_2Z, mRNA sequence.
ACCESSION AU031475
VERSION AU031475.1 GI:3767365
KEYWORDS EST.
SOURCE Oryza sativa.
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
1 (bases 1 to 510)
REFERENCE
AUTHORS Sasaki,T. and Yamamoto,K.
TITLE Rice cDNA from immature leaf including apical meristem
JOURNAL Unpublished (1997)
CONTACT: Takuji Sasaki
NATIONAL INSTITUTE OF AGROBIOLOGICAL RESOURCES
RICE GENOME RESEARCH PROGRAM
2-1-2 Kannondai,Tsukuba
Ibaraki,
Japan 305

```

Tel: 0298-38-7441
 Fax: 0298-38-7468
 Email: tsasaki@abr.affrc.go.jp
 PROJECT = 'RGP'

FEATURES

source Location/Qualifiers

1. 510
 /organism="Oryza sativa"
 /strain="Nipponbare"
 /db_xref="taxon:4530"
 /clone_1ib="E61659_22"
 /clone_1ib="Rice cDNA from immature leaf including apical meristem"
 /dev_stage="Immature"
 /note="Organ: Leaf; immature leaf including apical meristem (under long day condition)"

BASE COUNT 176 a 98 c 99 g 137 t

Query Match 64.0%; Score 16; DB 107; Length 510;
 Best Local Similarity 100.0%; Pred. No. 19;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 catgccagcattcg 23
 ||||||||||||
 Db 136 CATGCCAGCATTTG 121

RESULT 6

p968R 548 bp DNA GSS 25-JUL-2000
 LOCUS Leishmania major Friedlin PAC p968 right end-sequence, genomic survey sequence.
 DEFINITION AL390674
 VERSION AL390674.1 GI:9501650
 KEYWORDS GSS.
 SOURCE Leishmania major.
 ORGANISM Leishmania major.
 Eukaryota; Eukaryota; Kinetoplastida; Trypanosomatidae; Leishmania.

REFERENCE 1 (bases 1 to 548)
 Ivens, A.C., Lewis, S.M., Bagherzadeh, A., Zhang, L., Chan, H.M. and Smith, D.F.
 A physical map of the Leishmania major Friedlin genome
 Genome Res. 8 (2), 135-145 (1998)
 2 (bases 1 to 548)
 Taylor, R.G., Huckle, E.E.J., Ivens, A.C., Rajandream, M.A. and Barrell, B.G.
 Direct Submission
 Submitted (24-JUL-2000) Leishmania major Friedlin genome sequencing project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and allcates@sanger.ac.uk
 see http://www.ebi.ac.uk/parasites/leish.html
 Details of Leishmania sequencing at the Sanger Centre are available at http://www.sanger.ac.uk/projects/L_majior/
 The primer sequence can be obtained from alicat@sanger.ac.uk.

TITLE A physical map of the Leishmania major Friedlin genome
 JOURNAL Genome Res. 8 (2), 135-145 (1998)
 MEDLINE 98146435
 REFERENCE 2 (bases 1 to 548)
 Taylor, R.G., Huckle, E.E.J., Ivens, A.C., Rajandream, M.A. and Barrell, B.G.
 Direct Submission
 Submitted (24-JUL-2000) Leishmania major Friedlin genome sequencing project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and allcates@sanger.ac.uk
 see http://www.ebi.ac.uk/parasites/leish.html
 Details of Leishmania sequencing at the Sanger Centre are available at http://www.sanger.ac.uk/projects/L_majior/
 The primer sequence can be obtained from alicat@sanger.ac.uk.

TITLE A physical map of the Leishmania major Friedlin genome
 JOURNAL Genome Res. 8 (2), 135-145 (1998)
 MEDLINE 98146435
 REFERENCE 2 (bases 1 to 548)
 Taylor, R.G., Huckle, E.E.J., Ivens, A.C., Rajandream, M.A. and Barrell, B.G.
 Direct Submission
 Submitted (24-JUL-2000) Leishmania major Friedlin genome sequencing project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and allcates@sanger.ac.uk
 see http://www.ebi.ac.uk/parasites/leish.html
 Details of Leishmania sequencing at the Sanger Centre are available at http://www.sanger.ac.uk/projects/L_majior/
 The primer sequence can be obtained from alicat@sanger.ac.uk.

COMMENT

see http://www.ebi.ac.uk/parasites/leish.html
 Details of Leishmania sequencing at the Sanger Centre are available at http://www.sanger.ac.uk/projects/L_majior/
 The primer sequence can be obtained from alicat@sanger.ac.uk.

FEATURES
 source Location/Qualifiers

1. 548
 /organism="Leishmania major"
 /strain="Friedlin"
 /db_xref="taxon:5664"
 /clone="PAC p968"

BASE COUNT 86 a 167 c 163 g 132 t

Query Match 64.0%; Score 16; DB 258; Length 548;
 Best Local Similarity 100.0%; Pred. No. 19;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 5 gtgcgcagcagcatt 20

Db 541 GTGCATGCCAGCATTTG 526

RESULT 7
 AU030895/c 666 bp mRNA EST 29-OCT-1998
 LOCUS AU030895 Rice cDNA from immature leaf including apical meristem
 DEFINITION AU030895 Rice cDNA from immature leaf including apical meristem
 Oryza sativa cDNA clone E60405_22, mRNA sequence.

ACCESSION AU030895
 VERSION AU030895.1 GI:3766785
 KEYWORDS EST.
 SOURCE Oryza sativa.
 ORGANISM Oryza sativa.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE 1 (bases 1 to 666)
 Sasaki, T. and Yamamoto, K.
 Rice cDNA from immature leaf including apical meristem
 Unpublished (1997)
 Contact: Takuji Sasaki
 National Institute of Agrobiological Resources
 Rice Genome Research Program
 2-1-2 Kannondai, Tsukuba
 Ibaraki,
 Japan 305
 Tel: 0298-38-7441
 Fax: 0298-38-7468
 Email: tsasaki@abr.affrc.go.jp
 PROJECT = 'RGP'.

FEATURES
 source Location/Qualifiers

1. 666
 /organism="Oryza sativa"
 /strain="Nipponbare"
 /db_xref="taxon:4530"
 /clone="E60405_22"
 /clone_1ib="Rice cDNA from immature leaf including apical meristem"
 /dev_stage="Immature"
 /note="Organ: Leaf; immature leaf including apical meristem (under long day condition)"

BASE COUNT 195 a 149 c 131 g 181 t 10 others

Query Match 64.0%; Score 16; DB 107; Length 666;
 Best Local Similarity 100.0%; Pred. No. 19;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 catgccagcattcg 23
 ||||||||||||
 Db 293 CATGCCAGCATTTG 278

RESULT 8

AU078750 704 bp mRNA EST 18-OCT-1999
 LOCUS AU078750 Rice panicle at flowering stage Oryza sativa cDNA clone
 DEFINITION AU078750 Rice panicle at flowering stage Oryza sativa cDNA clone
 E0723_42, mRNA sequence.

ACCESSION AU078750
 VERSION AU078750.1 GI:6062509
 KEYWORDS EST.
 SOURCE Oryza sativa.
 ORGANISM Oryza sativa.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE 1 (bases 1 to 704)
 Sasaki, T. and Yamamoto, K.
 Rice cDNA from panicle at flowering stage
 Unpublished (1996)
 Contact: Takuji Sasaki

National Institute of Agrobiological Resources
Rice Genome Research Program
2-1-2 Kannondai, Tsukuba
Ibaraki,
Japan 305

Tel: 0298-38-7441
Fax: 0298-38-7468
Email: tsasakl@abr.affrc.go.jp

PROJECT = "RGP".

FEATURES

Source

Location/Qualifiers
1..704
/organism="Oryza sativa"
/strain="Nipponbare"
/db_xref="taxon:4530"
/clone="E0723.42"
/clone_lib="Rice panicle at flowering stage"
/dev_stage="flowering stage"
/note="Organ: panicle; Rice cDNA from panicle at flowering stage"
BASE COUNT 203 a 167 c 142 g 191 t 1 others
ORIGIN

Query Match 64.0%; Score 16; DB 107; Length 704;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 catgccagcttcg 23
|||||
Db 346 CATGCCAGCATTTG 331

RESULT 9
LOCUS BF864751 761 bp mRNA EST 19-JAN-2001
DEFINITION 963054E05.y1 C. reinhardtii CC-1690, stress condition I, normalized
ACCESSION BF864751
VERSION BF864751.1 GI:12254895
KEYWORDS EST.
SOURCE Chlamydomonas reinhardtii.
ORGANISM Chlamydomonas reinhardtii
Chlamydomonadaceae; Chlamydomonas.

REFERENCE 1 (bases 1 to 761)
Bukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;

AUTHORS Grossman, A., Davies, J., Federspiel, N., Harris, E., Hauser, C.,
Lefebvre, P., McDermott, J.P., Shlager, J., Sillfow, C. and Stern, D.

TITLE Analyses of the Chlamydomonas reinhardtii Genome: A Model,
Unicellular System for Analyzing Gene Function and Regulation in
Vascular Plants; project phase 3
JOURNAL Unpublished (2000)

COMMENT Contact: Charles Hauser
DCMB Box 91000
Duke University
Durham, NC 27708-1000

Tel: 919 613 8159
Fax: 919 613 8177
Email: chauser@duke.edu

FEATURES

Source

Location/Qualifiers
1..761
/organism="Chlamydomonas reinhardtii"
/strain="CC-1690 wild type mt+ 21gr"
/db_xref="taxon:3055"
/clone_lib="C. reinhardtii CC-1690, stress condition I,
normalized, lambda Zap II"
/note="Vector: pBluescript II SK-; Site,1: EcoRI; Site,2:
XhoI; This library, constructed by John Davies and Jeffrey
McDermott, combines cDNAs from CC-1690 cells grown to
mid-log phase in TAP-N (30 min, 1hr, 4hr), TAP-S (30 min,
1hr, 4hr), TAP-P (4hr, 12hr, 24hr), NO3 to NH4 (30min, 1hr,
4hr) and NH4 to NO3 (30min, 1hr, 4hr). PolyA mRNA was
purified from each sample, pooled and cDNA synthesized.
The cDNA was directionally cloned into lambda Zap II

(Stratagene) in the EcoRI (5') and XhoI (3') sites.
pBluescript II SK- plasmids were excised from the lambda
ZAP clones by superinfection with ExAssist (Stratagene)
phage. The library was normalized using method 4 described
in Bonaldo et al (1996) Genome Research 6: 791-806."

BASE COUNT 153 a 179 c 308 g 117 t 4 others
ORIGIN

Query Match 64.0%; Score 16; DB 170; Length 761;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 aggtgcacgcacga 18
|||||
Db 466 AGGTGCATGCCAGCA 451

RESULT 10
LOCUS A2578088 258 bp DNA GSS 08-DEC-2000
DEFINITION 18f02 Shot-gun genomic library of Rhizobium strain ANU265 Rhizobium
sp. NGR234 genomic clone 18f02, DNA sequence.
ACCESSION A2578088
VERSION A2578088.1 GI:11605015
KEYWORDS GSS.
SOURCE Rhizobium sp. NGR234.
ORGANISM Rhizobium sp. NGR234
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
Rhizobiaceae; Rhizobium.

REFERENCE 1 (bases 1 to 258)
Viprey, V., Kosenchal, A., Broughton, W.J. and Perret, X.
Genetic snapshots of the Rhizobium species NGR234 genome
Genome Biology.com 1 (6), 0014.1-0014.7 (2000)
COMMENT Laboratoire de Biologie Molculaire des Plantes Supérieures
University of Geneva
1 Chemin de l'Imperatrice, Chambesey/Geneva 1292, Switzerland
Tel: +44(0)1603450000
Fax: +44(0)1603450045
Email: virginie.viprey@bsrc.ac.uk
Class: shotgun.

FEATURES

Source

Location/Qualifiers
1..258
/organism="Rhizobium sp. NGR234"
/strain="ANU265"
/db_xref="taxon:394"
/clone="18f02"
/clone_lib="Shot-gun genomic library of Rhizobium strain
ANU265"
/note="Vector: M13; derivative strain of NGR234 cured of
pNGR234a"
BASE COUNT 51 a 82 c 79 g 46 t
ORIGIN

Query Match 60.0%; Score 15; DB 246; Length 258;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 tcgccagcttcga 24
|||||
Db 175 TCGCCAGCATTTGCA 161

RESULT 11
LOCUS A1410504 266 bp mRNA EST 09-FEB-1999
DEFINITION EST238797 Normalized rat heart, Bento Soares Rattus sp. cDNA clone
RHCCX29 3' end, mRNA sequence.
ACCESSION A1410504
VERSION A1410504.1 GI:4254008
KEYWORDS EST.

SOURCE Rattus sp.
ORGANISM Rattus sp.
REFERENCE 1 (bases 1 to 266)
AUTHORS Lee, N.H., Glodok, A., Chandra, I., Mason, T.M., Quackenbush, J., Kerlavage, A.R. and Adams, M.D.
TITLE Rat Genome Project: Generation of a Rat EST (REEST) Catalog & Rat Gene Index
JOURNAL Unpublished (1998)
COMMENT Contact: Lee, NH
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
Seq primer: M13-21.
FEATURES
Location/Qualifiers
1..266
/organism="Rattus sp."
/db_xref="taxon:10118"
/clone_lib="RHECX29"
/note="Organ: heart; Vector: pT7T3Pac; Site_1: EcoRI; Site_2: NotI"
BASE COUNT 56 a 61 c 64 g 85 t
ORIGIN
Query Match 60.0%; Score 15; DB 19; Length 266;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 4 ggtgcatgcgcagca 18
|||||
Db 90 ggtgcatgcgcagca 104
RESULT 12
BE065518/c 338 bp mRNA EST 09-JUN-2000
LOCUS
DEFINITION RC3-BT0316-170200-014-b06 BT0316 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE065518
VERSION BE065518.1 GI:8410168
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 338)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Negai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
2020263
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2=RC3-BT0316-170200-014-b06&t3=2000-02-17&t4=1)

Seq primer: puc 18 forward
High quality sequence start: 17
High quality sequence stop: 338.
FEATURES
source
1..338
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="BT0316"
/dev_stage="Adult"
/note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT 76 a 100 c 66 g 96 t
ORIGIN
Query Match 60.0%; Score 15; DB 162; Length 338;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 8 catgcgcagcatttc 22
|||||
Db 97 catgcgcagcatttc 83
RESULT 13
AV551712 341 bp mRNA EST 06-SEP-2000
LOCUS
DEFINITION AV551712 Arabidopsis thaliana roots Columbia Arabidopsis thaliana cDNA clone RZ13e05R 5', mRNA sequence.
ACCESSION AV551712
VERSION AV551712.1 GI:8723125
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 341)
Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.
A large scale analysis of cDNA in Arabidopsis thaliana: Generation of 12,028 non-redundant expressed sequence tags from normalized and size-selected cDNA libraries
DNA Res. 7, 175-180 (2000)
20363093
Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
FEATURES
source
1..341
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone_lib="RZ13e05R"
/clone_lib="Arabidopsis thaliana roots Columbia"
/tissue_type="roots"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT 73 a 85 c 53 g 130 t
ORIGIN
Query Match 60.0%; Score 15; DB 31; Length 341;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 10 tcgcgcagcatttcga 24

Insert Length: 1146 Std Error: 0.00
Seq primer: M13Rpl
High quality sequence stop: 180.

FEATURES

source

Location/Qualifiers
1. .394

/organism="Homo sapiens"
/db_xref="GDB:468473"
/db_xref="taxon:9606"
/clone="IMAGE:112856"
/clone_lib="Soares fetal liver spleen 1NPLS"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: liver and spleen; Vector: pT73D (Pharmacia) with a modified polylinker; Site_1: Pac I; Site_2: Eco RI; 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5' AACTGGAAGAAATTAATTAAGATCTTTTCTTTTCTTTT 3'] double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT
ORIGIN
86 a 98 c 82 g 122 t 6 others

Query Match 60.0%; Score 15; DB 169; Length 394;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 catgccagcatttc 22
Db 183 CATGCCAGCATTTTC 169

RESULT 17
BE429060/c 396 bp mRNA EST 26-JUN-2000
LOCUS MTD013.H06P990621 ITEC MTD Durum Wheat Root Library Trilicium
DEFINITION Trilicium subsp. durum cDNA clone MTD013.H06, mRNA sequence.
ACCESSION BE429060
VERSION BE429060.1 GI:9426903
KEYWORDS EST.
SOURCE durum wheat.
ORGANISM Trilicium turgidum subsp. durum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
; Triliceae; Trilicium.
1 (bases 1 to 396)

REFERENCE
AUTHORS Anderson, O.A., Appels, R., Bailey, P., Blake, T., Close, T., Cloutier, S., Dubcovsky, J., Feuillet, C., Gale, M., Graner, A., Gustafson, P., Hermann, R.G., Holton, T., Jacquemin, J.M., Jia, J., Joudrier, P., Langridge, P., Lazo, G.R., Lin, J.J., McGuire, P., Ogihara, Y., Pechonin, N., Qualset, C., Schuch, W., Selvaraj, G., Shariflou, M., Sorrells, M., Warburton, M. and Wenzel, G.
International Triliceae EST Cooperative (ITEC): Production of Expressed Sequence Tags for Species of the Triliceae
Unpublished (2000)

JOURNAL COMMENT
Contact: Joudrier P
IRMA, unite de Biochimie et Biologie Moléculaire des Cereales
2, place VIALA, 34060 Montpellier cedex 01 FRANCE
Tel: 33 4 99 61 23 84
Fax: 33 4 99 61 23 48
Email: joudrier@enscm.inra.fr
International Triliceae EST Cooperative (ITEC)
http://wheat.pw.usda.gov/genome.
Location/Qualifiers

FEATURES

source

1. .396
/organism="Trilicium turgidum subsp. durum"
/cultivar="Siliiana"
/db_xref="taxon:4567"
/clone="MTD013.H06"
/clone_lib="ITEC MTD Durum Wheat Root Library"

/tissue_type="root"
/dev_stage="3-day-old seedling, water-stressed"
/note="Vector: pSPORT1; T7 primers used. See pSPORT1 polylinker site. 0.3-2.0 kbp average insert size."

BASE COUNT
ORIGIN
96 a 87 c 116 g 97 t

Query Match 60.0%; Score 15; DB 167; Length 396;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcagtgatcgcaca 15
Db 75 GCAGTGATCGCACA 61

RESULT 18
BE044431 399 bp mRNA EST 08-JUN-2000
LOCUS h045d04.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
DEFINITION IMAGE:3040327 3', mRNA sequence.
ACCESSION BE044431
VERSION BE044431.1 GI:8361484
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 399)
NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40UP from Gibco.

FEATURES
source
1. .399
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3040327"
/clone_lib="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (fetal lung NBHL19W, testis NT7, and B-cell NCI CGAP GCBI) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 297480-302087, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT
ORIGIN
91 a 94 c 96 g 118 t

Query Match 60.0%; Score 15; DB 162; Length 399;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 catgccagcatttc 22
Db 329 CATGCCAGCATTTTC 343

RESULT 19
A1268648

LOCUS AI268648 414 bp mRNA EST 17-NOV-1998
 DEFINITION q039f05.x1 NCI_CGAP_Lu5 Homo sapiens cDNA clone IMAGE:1910913 3',
 mRNA sequence.
 ACCESSION AI268648
 VERSION AI268648.1 GI:3887815
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 414)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaaps-remail.nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LINL at:
www-bio.linnl.gov/bbrp/image/image.html
 Seq primer: -400P from Gibco
 High quality sequence stop: 398.
 Location/Qualifiers
 1..414
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1910913"
 /clone_1id="NCI_CGAP_Lu5"
 /tissue_type="carcinoid"
 /lab_host="DH10B"
 /note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a
 modified polylinker; 1st strand cDNA was prepared from a
 neuroendocrine lung carcinoid, and was then primed with a
 Not I - oligo(dT) primer. Double-stranded cDNA was ligated
 to Eco RI adaptors (Pharmacia), digested with Not I and
 cloned into the Not I and Eco RI sites of the modified
 pT73 vector. Library is normalized. Library was
 constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 122 a 82 c 109 g 101 t
 ORIGIN

Query Match 60.0%; Score 15; DB 18; Length 414;
 Best Local Similarity 100.0%; Pred. No. 72;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 catgccagcatttc 22
 |||||
 Db 353 CATGCCAGCATTTTC 367

RESULT 20
 LOCUS AQ027122 435 bp DNA GSS 30-JUN-1998
 DEFINITION CIT-HSP-2323H13.TF CIT-HSP Homo sapiens genomic clone 2323H13, DNA
 sequence.
 ACCESSION AQ027122
 VERSION AQ027122.1 GI:3267344
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 435)
 Adams, M.D., Rounsley, S.D., Zhao, S., Field, C.E., Bass, S., Linher, K.,
 Golden, K., Berry, K., Granger, D., Suh, E., Wible, C., Shizuya, H.,
 Simon, M., and Venter, J.C.
 TITLE Use of a random BAC End Sequence Database for Sequence-Ready Map

JOURNAL Building (1998)
 COMMENT Unpublished (1998)
 Contact: Mark Adams
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: mdamad@tigr.org
 Clones are available from Research Genetics (info@resgen.com). BAC
 end search page:
http://www.tigr.org/cdb/humgen/bac_end_search/bac_end_search.html.
 Seq primer: M13-21
 Class: BAC ends.
 Location/Qualifiers
 1..435
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="2323H13"
 /clone_1id="CIT-HSP"
 /sex="Male"
 /cell_type="Sperm"
 /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
 HindIII"

BASE COUNT 140 a 97 c 77 g 121 t
 ORIGIN

Query Match 60.0%; Score 15; DB 223; Length 435;
 Best Local Similarity 100.0%; Pred. No. 72;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 catgccagcatttc 22
 |||||
 Db 178 CATGCCAGCATTTTC 192

RESULT 21
 LOCUS H69200/c 437 bp mRNA EST 19-OCT-1995
 DEFINITION YR59f10.r1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone
 IMAGE:213067 5', mRNA sequence.
 ACCESSION H69200
 VERSION H69200.1 GI:1030526
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 437)
 Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,
 Chissole, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W., Hawkins
 , M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Mardis, E., Moore
 , B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T.,
 Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevas, E.,
 Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Marra, M.
 Generation and analysis of 280,000 human expressed sequence tags
 Genome Res. 6 (9), 807-828 (1996)
 97044478
 CONTACT Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 Insert Size: 1098
 High quality sequence stops: 300
 Source: IMAGE Consortium, LINL
 This clone is available royalty-free through LINL; contact the
 IMAGE Consortium (info@image.linnl.gov) for further information.
 Insert length: 1098 Std Error: 0.00
 Seq primer: M13RP1
 High quality sequence stop: 300.

FEATURES
source

Location/Qualifiers

1. .437
/organism="Homo sapiens"
/db_xref="GDB:3778371"
/db_xref="taxon:9606"
/clone_image="IMAGE:213067"
/clone_lib="Soares fetal liver spleen INFLS"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: pTT3D (Pharmacia) with a modified polylinker. Site_1: Pac I; Site_2: Eco RI; 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5' AACTGACAGATTATTAAGAATCTTTTCTTTTCTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pTT3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT
ORIGIN

97 a 113 c 89 g 134 t 4 others

Query Match 60.0%; Score 15; DB 158; Length 437;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 catgccagcatc 22
|||||
DB 139 CATGCCAGCATTC 125

RESULT 22
N92081 438 bp mRNA EST 04-APR-1996
LOCUS z23h03.r1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone
DEFINITION IMAGE:293429 5', mRNA sequence.

ACCESSION N92081
VERSION N92081.1 GI:1264390
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 438)
Mammalia: Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman
M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J.,
Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston
R., Williamson, A., Woldmann, P. and Wilson, R.

TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

FEATURES
source
1. .438
/organism="Homo sapiens"
/db_xref="GDB:3801206"
/db_xref="taxon:9606"
/clone_image="IMAGE:293429"
/clone_lib="Soares fetal liver spleen INFLS"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: pTT3D (Pharmacia) with a modified polylinker. Site_1: Pac I; Site_2: Eco RI;
High quality sequence stop: 321.
Location/Qualifiers

1. .438
/organism="Homo sapiens"
/db_xref="GDB:3801206"
/db_xref="taxon:9606"
/clone_image="IMAGE:293429"
/clone_lib="Soares fetal liver spleen INFLS"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: pTT3D (Pharmacia) with a modified polylinker. Site_1: Pac I; Site_2: Eco RI;

1st strand cDNA was primed with a Pac I - oligo(dT) primer [5' AACTGACAGATTATTAAGAATCTTTTCTTTTCTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pTT3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT
ORIGIN

104 a 111 c 101 g 118 t 4 others

Query Match 60.0%; Score 15; DB 187; Length 438;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 catgccagcatc 22
|||||
DB 308 CATGCCAGCATTC 294

RESULT 23
AV722315 445 bp mRNA EST 16-OCT-2000
LOCUS AV722315
DEFINITION AV722315 HTB Homo sapiens cDNA clone HTBAMG01 5', mRNA sequence.
ACCESSION AV722315
VERSION AV722315.1 GI:10824678
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 445)
Mammalia: Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
Gu, Y., Peng, Y., Song, H., Huang, Q., Yang, Y., Gao, G., Xiao, H., Xu, X.,
Li, N., Qian, B., Liu, F., Qu, J., Gao, X., Cheng, Z., Xu, Z., Zeng, L., Xu
S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu, G., Hu, R.,
Chen, J., Chen, Z. and Han, Z.
Homo sapiens cDNA HTB clones
Unpublished (2000)
Contact: Zeguang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhanjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919 (ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.

FEATURES
source

Location/Qualifiers

1. .445
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="HTBAMG01"
/clone_lib="HTB"
/tissue="Hypothalamus"
/dev_stage="Adult"
/lab_host="SOLR"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT 102 a 123 c 104 g 116 t
ORIGIN

Query Match 60.0%; Score 15; DB 32; Length 445;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 catgccagcatc 22
|||||

DB 296 CATGCCAGCATTC 282

RESULT 24
A0166733/c 453 bp mRNA EST 23-JAN-2001
LOCUS A0166733

DEFINITION AU166733 Rice callus *Oryza sativa* subsp. *japonica* cDNA clone C50066
, mRNA sequence.
ACCESSION AU166733
VERSION AU166733.1 GI:12405132
KEYWORDS EST.
SOURCE *Oryza sativa* subsp. *japonica*.
ORGANISM *Oryza sativa* subsp. *japonica*.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; *Oryza*.
REFERENCE 1 (bases 1 to 453)
AUTHORS Sasaki, T. and Yamamoto, K.
TITLE Rice cDNA from callus (2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program
2-1-2 Kannondai, Tsukuba
Ibaraki,
Japan 305
Tel: 0298-38-7441
Fax: 0298-38-7468
Email: tsasaki@abr.affrc.go.jp
PROJECT = 'RGP'.
C50066_98Z.

FEATURES
source
1..453
location/Qualifiers
/organism="Oryza sativa subsp. japonica"
/strain="cultivar Nipponbare, sub-species japonica"
/db_xref="taxon:39947"
/clone="C50066"
/clone_lib="Rice callus"
/note="Vector: pBluescript II SK+; Site_1: SalI; Site_2:
NotI; cDNA prepared from rice callus mRNAs by using
oligo(dT) as a primer and ligating to the SalI-NotI site
of pBluescript II SK+ phagemid."
BASE COUNT 164 a 83 c 140 g 66 t
ORIGIN

Query Match 60.0%; Score 15; DB 108; Length 453;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 tgcacgcacgacat 20
|||||
Db 349 TGCATGCCACGACAT 335

RESULT 25
BE446670/c 463 bp mRNA EST 25-JUL-2000
LOCUS WHE1139_F07_L13Z5 Wheat etiolated seedling root normalized cDNA
DEFINITION library Triticum aestivum cDNA clone WHE1139_F07_L13, mRNA
sequence.
ACCESSION BE446670
VERSION BE446670.1 GI:9446232
KEYWORDS EST.
SOURCE bread wheat.
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 463)
AUTHORS Anderson, O.D., Chao, S., Choi, D.W., Close, T.J., Fenton, R.D., Han,
P.S., Hsia, C.C., Kang, Y., Lazo, G.R., Miller, R., Nguyen, H.T.,
Rausch, C.J., Seaton, C.L., Tong, J.C. and Zhang, D.
TITLE The structure and function of the expressed portion of the wheat
genomes - Normalized root cDNA library
JOURNAL Unpublished (2000)
COMMENT Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific
West Area, Western Regional Research Center

800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105595818
Email: oanderson@pw.usda.gov
Sequence have been trimmed to remove vector sequence and low
quality sequence with phred score less than 20
Seq primer: Stragene SK primer.
location/Qualifiers
1..463
/organism="Triticum aestivum"
/cultivar="Chinese Spring"
/db_xref="taxon:4565"
/clone="WHE1139_F07_L13"
/clone_lib="Wheat etiolated seedling root normalized cDNA
library"
/tissue_type="Root"
/dev_stage="Five day old etiolated seedling"
/lab_host="E. coli DH10b"
/note="Vector: lambda Uni-ZAP XR, excised phagemid
pBluescript SK; Site_1: EcoRI; Site_2: XhoI; Seeds were
surface-sterilized, germinated and grown aseptically in
the dark at room temperature on filter paper with water,
nystatin and cefotaxime in covered crystallization
dishes. Roots were harvested. The tissue, total RNA, and
poly(A) RNA were prepared, a cDNA library was made in the
TJ Close lab (Choi, Close, Fenton) at the University of
California, Riverside. The cDNA clones were in vivo
excised to give pBluescript phagemids before
normalization was carried out. The mass excision of
phagemid library and normalization were done in HT Nguyen
lab by D. Zhang at Texas Tech University. Normalization
protocol used was that of Soares. Plasmid DNA
preparations and DNA sequencing were performed in the OD
Anderson lab (all other authors)."
BASE COUNT 119 a 98 c 132 g 114 t
ORIGIN

Query Match 60.0%; Score 15; DB 167; Length 463;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 gcaagtgcatgccca 15
|||||
Db 111 GCAGTGCATGCCCA 97

RESULT 26
AA026650/c 466 bp mRNA EST 09-MAY-1997
LOCUS 2399e08_r1 Soares_pregnant_uterus_NbHFO Homo sapiens cDNA clone
DEFINITION IMAGE:469190 5', mRNA sequence.
ACCESSION AA026650
VERSION AA026650.1 GI:1492715
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 466)
AUTHORS Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,
Chissoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W., Hawkins,
M., Hultman, M., Kucaba, T., Lacy, M., Le, N., Mardis, E., Moore,
B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T.,
Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevaskis, E.,
Underwood, K., Wondolmnn, F., Waterston, R., Wilson, R. and Marra, M.
TITLE Generation and analysis of 280,000 human expressed sequence tags
JOURNAL Genome Res. 6 (9), 807-828 (1996)
MEDLINE 97044478
COMMENT Contact: Willson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800

Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 Insert Length: 1129 Std Error: 0.00
 Seq primer: -28M13 rev2 from Amersham
 High quality sequence stop: 364.
 Location/Qualifiers

FEATURES

source
 1..466
 /organism="Homo sapiens"
 /db_xref="GDB:3755192"
 /db_xref="taxon:9606"
 /clone="IMAGE:469190"
 /clone_lib="Soares_pregnant_uterus_NBHPV"
 /sex="female"
 /dev_stage="adult"
 /lab_host="DH10B"
 /note="Organ: uterus; Vector: pT73-Pac; Site: 1: Not I;
 Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
 oligo(dT) primer [5',
 AACTGAGAGATTCGCGCGCCCTTTTCTTTTCTTTT 3'],
 double-stranded cDNA was ligated to Eco RI adaptors
 (pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pT73 vector. Library
 went through one round of normalization. Library
 constructed by M. Fatima Bonaldo."
 BASE COUNT 114 a 124 c 100 g 128 t
 ORIGIN

Query Match 60.0%; Score 15; DB 1; Length 466;
 Best Local Similarity 100.0%; Pred. No. 72;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 8 catgccagcattc 22
 |||||
 Db 263 catgccagcatttc 249

RESULT 27
 AA821080 475 bp mRNA EST 25-NOV-1998
 LOCUS
 DEFINITION GM096622 5prime GM Drosophila melanogaster ovary Bluescript
 Drosophila melanogaster CDNA clone GM096622 5prime, mRNA sequence.
 ACCESSION AA821080
 VERSION AA821080.1 GI:2890948
 KEYWORDS EST.
 SOURCE
 ORGANISM
 Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 475)
 Harvey, D., Hong, L., Evans-Holm, M., Pendleton, J., Su, C., Brokstein
 P., Lewis, S. and Rubin, G.M.
 BDCP/HHMI Drosophila EST Project
 Unpublished (1997)
 Contact: Harvey, D.
 G. M. Rubin-Molecular and Cell Biology
 University of California Berkeley
 539 USA, Berkeley, CA 94720-3200, USA
 Fax: 510 643 9947
 Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
 Plate: 96 row: B column: 10
 High quality sequence stop: 439.
 Location/Qualifiers

FEATURES

source
 1..475
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone="GM096622"
 /clone_lib="GM Drosophila melanogaster ovary Bluescript"
 /sex="female"
 /dev_stage="newly eclosed females; germlarium-stage 6"

/lab_host="SOLR"
 /note="Organ: ovary; Vector: Bluescript SK; Site: 1: EcoRI;
 Site_2: XhoI; Constructed using Stratagene ZAP-cDNA
 Synthesis kit. Oligo dt-primed and directionally cloned at
 EcoRI and XhoI in Bluescript SK(+/-)"
 BASE COUNT 128 a 147 c 121 g 79 t
 ORIGIN

Query Match 60.0%; Score 15; DB 12; Length 475;
 Best Local Similarity 100.0%; Pred. No. 72;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 3 aggtgcagccagc 17
 |||||
 Db 344 AGGTGCATCCGACG 358

RESULT 28
 H60979 475 bp mRNA EST 06-OCT-1995
 LOCUS
 DEFINITION Y122F01.r1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone
 IMAGE:206041 5', mRNA sequence.
 ACCESSION H60979
 VERSION H60979.1 GI:1013811
 KEYWORDS EST.
 SOURCE
 ORGANISM
 human.
 Homo sapiens
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 475)
 Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman
 M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J.,
 Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevasakis, E., Waterston
 R., Williamson, A., Woldmann, P. and Wilson, R.
 The WashU-Merck EST Project
 Unpublished (1995)
 Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 Insert Size: 1101
 High quality sequence stops: 339
 Source: IMAGE Consortium, LNL
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 Insert Length: 1101 Std Error: 0.00
 Seq primer: M13RPI
 High quality sequence stop: 339.
 Location/Qualifiers

FEATURES

source
 1..475
 /organism="Homo sapiens"
 /db_xref="GDB:3775172"
 /db_xref="taxon:9606"
 /clone="IMAGE:206041"
 /clone_lib="Soares fetal liver spleen INFLS"
 /sex="male"
 /dev_stage="20 week post conception fetus"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia)
 with a modified polylinker; Site: 1: Pac I; Site_2: Eco RI;
 1st strand cDNA was primed with a Pac I - oligo(dT) primer
 [5', AACTGAGAGATTAATTAAGATCTTTTCTTTTCTTTT 3'],
 double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Pac I and cloned into the Pac I
 and Eco RI sites of the modified pT73 vector. Library
 went through one round of normalization. Library
 constructed by Bento Soares and M. Fatima Bonaldo."
 BASE COUNT 109 a 123 c 95 g 143 t 5 others
 ORIGIN

Query Match 60.0%; Score 15; DB 158; Length 475;
 Best Local Similarity 100.0%; Pred. No. 72;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 catgccagcatttc 22
 |||||||

DB 119 CATGCCAGCATTTC 105

RESULT 29
 BG101817 475 bp mRNA EST 30-JAN-2001
 LOCUS RH122.22.G09.g1.A003 Rhizome2 (RH122) Sorghum prolinguam cDNA, mRNA
 DEFINITION
 ACCESSION BG101817
 VERSION BG101817.1 GI:12616762
 KEYWORDS EST.
 SOURCE Sorghum prolinguam.
 ORGANISM Sorghum prolinguam.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade; Panicoideae; Andropogoneae; Sorghum.
 I (bases 1 to 475)
 Cordonnier-Pratt, M.-M., Gingle, A., Paterson, A., Sudman, M. and Pratt
 , L.H.
 TITLE An EST database from Sorghum: Sorghum prolinguam rhizomes
 JOURNAL Unpublished (2000)
 COMMENT Contact: Cordonnier-Pratt MM
 Department of Botany
 The University of Georgia
 Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
 Tel: 706 542 1860
 Fax: 706 542 1805
 Email: mmp@pratt.uga.edu
 Seq primer: POLYTMix
 High quality sequence start: 11
 High quality sequence stop: 408
 POLYA-No.

FEATURES
 source location/Qualifiers
 1..475
 /organism="Sorghum prolinguam"
 /db_xref="taxon:132711"
 /clone_lib="Rhizome2 (RH122)"
 /note="Organ: Rhizomes; Vector: pBluescript II from lambda
 zap II; Site 1: XhoI; Site 2: EcoRI; The library was made
 from poly-A RNA in the cloning vector lambda zap II.
 Clones to be sequenced were prepared by mass excision."
 BASE COUNT 128 a 107 c 113 g 127 t

ORIGIN
 1 60.0%; Score 15; DB 173; Length 475;
 Best Local Similarity 100.0%; Pred. No. 72;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ggcggatgcacgccca 15
 |||||||

DB 240 GCAGGTGCATGCCCA 254

RESULT 30
 AI692369 491 bp mRNA EST 17-DEC-1999
 LOCUS wd63f02.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:2336283 3',
 DEFINITION mRNA sequence.
 ACCESSION AI692369
 VERSION AI692369.1 GI:4969709
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 491)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabs-r@mail.nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 www.bio.llnl.gov/bbrp/image/image.html
 Insert length: 569 Std Error: 0.00
 Seq primer: 40UP from Gibco
 High quality sequence stop: 471.
 Location/Qualifiers
 1..491
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="IMAGE:2336283"
 /clone_lib="NCI-CGAP_Lu24"
 /rname_type="carcinoid"
 /lab_host="DH10B"
 /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
 modified polylinker; Plasmid DNA from the normalized
 library NCI-CGAP Lu5 was prepared and ss circles were
 made in vitro. Following HAP purification, this DNA was
 used as tracer in a subtractive hybridization reaction.
 The driver was PCR-amplified cDNAs from a pool of 5,000
 clones made from the same library (clones
 141920-1417991 and 1520904-1522439). Subtraction by Bento
 Soares and M. Fatima Bonaldo."
 BASE COUNT 151 a 94 c 129 g 116 t 1 others

ORIGIN
 1 60.0%; Score 15; DB 23; Length 491;
 Best Local Similarity 100.0%; Pred. No. 72;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 catgccagcatttc 22
 |||||||

DB 344 CATGCCAGCATTTC 358

RESULT 31
 BF211097 491 bp mRNA EST 06-NOV-2000
 LOCUS 601812706P1 NIH-MGC_54 Homo sapiens cDNA clone IMAGE:4047120 5',
 DEFINITION mRNA sequence.
 ACCESSION BF211097
 VERSION BF211097.1 GI:11104683
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 491)
 NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabs-r@mail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: CLONETECH Laboratories, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>

Plate: L1CM875 row: 1 column: 01
 High quality sequence start: 56
 High quality sequence stop: 491.
 Location/Qualifiers

FEATURES

source

1.491
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="IMAGE:4047120"
 /clone_lib="NIH.MGC.54"
 /tissue_type="from chronic myelogenous leukemia"
 /lab_host="DH10B (TI phage-resistant)"
 /note="Organ: bone marrow; Vector: pDNR-LIB (Clontech);
 Site_1: SfiI (ggcgcgcgcgc); Site_2: SfiI (ggccatagcgc
); Double-stranded cDNA was prepared from cell line RNA.
 5' and 3' adaptors were used in cloning as follows: 5'
 adaptor sequence: 5'-CAGGCGCATTTATGCGC-3' and 3' adaptor
 sequence: 5'-ATCTAGAGCGCGAGCGCGCATTCG-dT(30)BN-3'
 (where B = A, C, G, or T and N = A, C, G, or T). Average
 insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies
 contained inserts by PCR. This library was enriched for
 full-length clones and was constructed by Clontech
 Laboratories (Palo Alto, CA)."
 BASE COUNT 123 a 128 c 119 g 121 t
 ORIGIN

Query Match 60.0%; Score 15; DB 146; Length 491;
 Best Local Similarity 100.0%; Pred. No. 72;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 catgcccagcatttc 22
 |||||

Db 305 CATTGCCAGCATTC 291

RESULT 32
 LOCUS BF916376 499 bp mRNA EST 18-JAN-2001
 DEFINITION IL3-UT0115-111200-377-F05 UT0115 Homo sapiens cDNA, mRNA sequence.
 ACCESSION BF916376
 VERSION BF916376.1 GI:12307834
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 499)
 REFERENCE
 AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
 Nagai,M.A., da Silva,W. Jr., Zagro,M.A., Bordin,S., Costa,F.F.,
 Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
 Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
 ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
 Simpson,A.J.
 Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

TITLE
 JOURNAL MEDLINE
 COMMENT 20202663
 Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?fl=IL3&cl=IL3-UT0115-
 11200-377-F05&ts=2000-12-11&la=1)
 Seq primer: puc 18 forward
 High quality sequence stop: 496.
 Location/Qualifiers
 1.499

FEATURES
 source

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="UT0115"
 /dev_stage="Adult"
 /note="Organ: uterus,tumor; Vector: puc18; Site_1: SmaI;
 Site_2: SmaI; A mini-library was made by cloning products
 derived from ORESTES PCR (U.S. Letters Patent application
 No. 196,716 - Ludwig Institute for Cancer Research)
 profiles into the pUC 18 vector. Reverse transcription of
 tissue mRNA and cDNA amplification were performed under
 low stringency conditions."
 BASE COUNT 121 a 121 c 114 g 143 t
 ORIGIN

Query Match 60.0%; Score 15; DB 171; Length 499;
 Best Local Similarity 100.0%; Pred. No. 72;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcaagtgcagccca 15
 |||||

Db 216 GCAGGTGCATGCCCA 230

RESULT 33
 LOCUS C98143 505 bp mRNA EST 29-APR-1999
 DEFINITION C98143 Rice callus Oryza sativa subsp. japonica cDNA clone C0777_7A
 , mRNA sequence.
 ACCESSION C98143 D28202
 VERSION C98143.1 GI:3760889
 KEYWORDS EST.
 SOURCE Oryza sativa subsp. japonica.
 ORGANISM Oryza sativa subsp. japonica.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Eriharoidae; Oryzaceae; Oryza.
 1 (bases 1 to 505)
 REFERENCE
 AUTHORS Yamamoto,K. and Sasaki,T.
 Rice cDNA from callus (1998)
 TITLE Unpublished (1998)
 JOURNAL On Oct 19, 1998 this sequence version replaced gi:454479.
 COMMENT National Institute of Agrobiological Resources
 Rice Genome Research Program
 2-1-2 Kannondai,Tsukuba
 Ibaraki,
 Japan 305
 Tel: 0298-38-7441
 Fax: 0298-38-7468
 Email: tsasaki@ab.affrc.go.jp
 D28202: Submitted (01-JAN-1994)
 PROJECT "RG"
 POLYA=No.
 Location/Qualifiers
 1.505
 /organism="Oryza sativa subsp. japonica"
 /strain="cultivar Nipponbare, sub-species japonica"
 /db_xref="taxon:39947"
 /clone_lib="C0777_7A"
 /clone_lib="Rice callus"
 /note="Vector: Bluescript II SK+; Site_1: SalI; Site_2:
 oligo(dT) as a primer and ligating to the SalI-NotI site
 of pBluescript II SK+ phagemid.
 2 others

FEATURES
 source

BASE COUNT 158 a 103 c 170 g 72 t
 ORIGIN

Query Match 60.0%; Score 15; DB 156; Length 505;
 Best Local Similarity 100.0%; Pred. No. 73;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy	6	tgcctccagcattt 20
Db	475	tgcctccagcattt 461
RESULT	34	
LOCUS	T84279/c	507 bp mRNA EST 16-MAR-1995
DEFINITION	yda7b04.r1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone IMAGE:111343 5', mRNA sequence.	
ACCESSION	T84279	
VERSION	T84279.1	GI:712567
KEYWORDS	EST.	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 507)	
AUTHORS	Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,R.M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevisan,E., Waterston,R., Williamson,A., Mohlmann,P. and Wilson,R. The WashU-Merck EST Project Unpublished (1995)	
TITLE	Contact: Wilson RK	
JOURNAL	Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel.: 314 286 1800 Fax: 314 286 1810 Email: est@wustl.wustl.edu	
COMMENT	Insert Size: 1225 High quality sequence stops: 283 Source: IMAGE Consortium, LNL This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.lnl.gov) for further information. Insert Length: 1225 Std Error: 0.00 Seq primer: M13Rp1 High quality sequence stop: 283. Location/Qualifiers 1..507 /organism="Homo sapiens" /db_xref="GDB:466960" /db_xref="taxon:9606" /clone="IMAGE:111343" /clone_lib="Soares fetal liver spleen INFLS" /sex="male" /dev_stage="20 week post conception fetus" /lab_host="DH10B (ampicillin resistant)" /note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia with a modified polylinker; Site_1: Pac I; Site_2: Eco RI; 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5' AACGTGAAGATTAATTAAAGACTTTTCTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."	
BASE COUNT	106 a 125 c 114 g 159 t	3 others
ORIGIN		
Query Match	60.0%; Score 15; DB 189; Length 507;	
Best Local Similarity	100.0%; Pred. No. 73;	
Matches	15; Conservative	0; Mismatches
Oy	8 catgccagacttc 22 	Indels
Db	192 catgccagcatttc 178	Gaps
RESULT	35	
LOCUS	BG467085	514 bp mRNA EST 20-MAR-2001
DEFINITION	1A04C06 Bovine Mixed Adipose cDNA library Bos taurus cDNA 5', mRNA	

sequence.
BG467085
GI:1396060

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Dixon, S.S., Hansen, C., Li, C., Fu, A., Meng, Y., Li, G., Murdoch, G.,
Moore, S.S., and Christopherson, B.
cDNA's from bovine subcutaneous adipose tissue
unpublished (2001)
Contact: Dr. Stephen Moore
. Beef Genomics Laboratory
Dept of AFNS, University of Alberta
410 Agri/Foer, Dept of AFNS, U of A, Edmonton, AB, T6G 2P5, Canada
Tel.: 780 492 0169
Fax: 780 492 4265
Email: smoores@afns.ualberta.ca
The sequence best matches gb:BTNDHDB (B.taurus CI-BI4.5b mRNA for
NADH dehydrogenase (ubiquinone)) in main database at high score of
908.0 and E-value of 0.0
PCR Primers
FORWARD: M13 Forward
BACKWARD: M13 Reverse
Seq primer: T3 primer
High quality sequence stop: 514
POLYA-No.

FEATURES
source
Location/Qualifiers
1..514
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="Bovine Mixed Adipose cDNA library"
/sex="two males and one female mixed"
/tissue_type="Adipose"
/cell_type="Adipocyte"
/dev_stage="Young adult"
/lab_host="XL1-BlueMRF'strain"
/note="Organ: Subcutaneous, Omental, Mesenteric adipose;
Vector: Uni-ZAPvXr, Site:1; EcoRI; Site_2: Xho I"

BASE COUNT
ORIGIN
115 a 112 c 132 g 155 t

Query Match 60.0%; Score 15; DB 154; Length 514;
Best Local Similarity 100.0%; Pred. No. 73;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 tgcacgccagcatc 20
|||||
|||||

Db 242 TGCAATGCCAGCATT 256

RESULT 36
LOCUS
AI957283/c 524 bp mRNA EST 20-AUG-1999
DEFINITION
ui85ei2.xl Sugano mouse kidney mkia Mus musculus cDNA clone
IMAGE:2158894 3', mRNA sequence.
ACCESSION
AI957283
AI957283.1 GI:5749992
VERSION
KEYWORDS
SOURCE
house mouse.
Mus musculus
Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 524)

REFERENCE
AUTHORS
Marras, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,
Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person,
B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter,
E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,
Westerston, R. and Wilson, R.

Query Match 60.0%; Score 15; DB 237; Length 545;
 Best Local Similarity 100.0%; Pred. No. 73;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcaggtgcacgcga 15
 |||
 DB 505 GCAGGTGCACGCCA 491

RESULT 41
 AM588900/c 547 bp mRNA EST 05-APR-2001
 LOCUS
 DEFINITION r06g05.y2 Bird-Rao Meloidogyne incognita J2 meloidogyne incognita
 cDNA 5', similar to TR:055013 055013 BRT3.; contains Alu repetitive
 element.; mRNA sequence.

ACCESSION AM588900.1 GI:7275932
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 EST.
 southern root-knot nematode.
 Meloidogyne incognita
 Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchima;
 Tylenchoidea; Heterodertiae; Meloidogyninae; Meloidogyne.

REFERENCE
 ADTHORS
 1 (bases 1 to 547)
 McCarter, J., Clifton, S., Chiapelli, B., Page, D., Martin, J., Wylie, T.,
 Dante, M., Marra, M., Hillier, L., Kucaba, T., Theising, B., Bowers, Y.,
 Gibbons, M., Ritter, E., Bennett, J., Franklin, C., Tsagaris, H., R.,
 Rook, L., Kennedy, S., Maguire, L., Beck, C., Underwood, K., Steptoe,
 M., Allen, M., Person, B., Swaller, T., Harvey, N., Schurk, R., Kohn, S.,
 Shurt, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and
 Wilson, R.

TITLE
 JOURNAL
 COMMENT
 The Washington Univ. Nematode EST Project, 1999
 Unpublished (1999)
 Contact: McCarter JP
 Washington Univ., Nematode EST Project, 1999
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810

Email: est@wustl.wustl.edu
 The library was constructed by Uma Rao and David Bird
 (david_bird@wustl.edu) at North Carolina State University. DNA
 Sequencing by: Washington University Genome Sequencing Center St.
 Louis.

Seq primer: T3 ET from Amersham
 High quality sequence stop: 519.

FEATURES
 source
 Location/Qualifiers
 1..547

/organism="Meloidogyne incognita"
 /db_xref="taxon:6306"
 /clone_lib="Bird-Rao Meloidogyne incognita J2"
 /dev_stage="enriched for 2nd stage juveniles"
 /lab_host="X10RL"

/note="Vector: ZAP express - PBKCMV (Stratagene); Site.1:
 EcoRI; Site.2: XhoI; Oligo (dT) primed library. cDNA was
 constructed and cloned unidirectionally into the vector
 within the 5' EcoRI and 3' XhoI sites. This library was
 constructed by Dr. Uma Rao and Dr. David Bird at North
 Carolina state University."

BASE COUNT 179 a 76 c 106 g 185 t 1 others
 ORIGIN

Query Match 60.0%; Score 15; DB 118; Length 547;
 Best Local Similarity 100.0%; Pred. No. 73;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 catgccagcattc 22
 |||
 DB 296 CATGCCAGCATTTC 282

RESULT 42
 A2283519/c 553 bp DNA GSS 27-JUN-2000
 LOCUS
 DEFINITION RPCI-23-125E16.TJ RPCI-23 Mus musculus genomic clone RPCI-23-125E16
 / DNA sequence.

ACCESSION A2283519.1 GI:9525226
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 house mouse.
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

REFERENCE
 ADTHORS
 1 (bases 1 to 553)
 Zhao, S., Nieman, W., Feldblyum, T., Malek, J., Shatsman, S., Akinet,
 B., Levins, M., Megann, S., Tsegaye, G., Geer, K., Kroll, M., de Jong, P.,
 and Fraser, C.M.
 Mouse BAC End Sequences from Library RPCI-23
 Unpublished (1999)

Other GSSs: RPCI-23-125E16.TV
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208

Email: szhao@tigr.org
 Clones are derived from the mouse BAC library RPCI-23. For BAC
 library availability, please contact Pieter de Jong
 (pdejong@med.buitalo.edu). Clones may be purchased from
 BACPAC Resources (<http://bacpac.med.buitalo.edu/orderingframe.htm>)
 or from Resea ch Genetics (info@resgen.com). BAC end page:
http://www.tigr.org/tdb/bac/ends/mouse/bac_end_intro.html
 Plate: 125 row: E column: 16

Seq primer: SP6
 Class: BAC ends.

FEATURES
 source
 Location/Qualifiers
 1..553

/organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPCI-23-125E16"
 /clone_lib="RPCI-23"
 /sex="Female"
 /lab_host="DH10B"

/note="Organ: Kidney/Brain; Vector: PBACe3.6; Site.1:
 EcoRI; Site.2: EcoRI; Female C57BL/6J mouse kidney and/or
 brain genomic DNA was isolated and partially digested
 with a combination of EcoRI and EcoRI methylase. Size
 selected DNA was cloned into the PBACe3.6 vector at the
 EcoRI sites. The ligation products were transformed into
 DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 135 a 156 c 126 g 136 t
 ORIGIN

Query Match 60.0%; Score 15; DB 241; Length 553;
 Best Local Similarity 100.0%; Pred. No. 73;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcaggtgcacgcga 15
 |||
 DB 525 GCAGGTGCACGCCA 511

RESULT 43
 C98142/c 567 bp mRNA EST 19-OCT-1998
 LOCUS
 DEFINITION C98142 Rice callus Oryza sativa subsp. japonica cDNA clone
 C0777.10Z, mRNA sequence.

ACCESSION C98142
 VERSION C98142.1 GI:3760888
 KEYWORDS
 SOURCE
 Oryza sativa subsp. japonica.

ORGANISM *Oryza sativa* subsp. *japonica*
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzoae; *Oryza*.
 REFERENCE 1 (bases 1 to 567)
 AUTHORS Sasaki,T. and Minobe,Y.
 TITLE Rice cDNA from callus
 JOURNAL Unpublished (1994)
 COMMENT Contact: Takuji Sasaki
 National Institute of Agrobiological Resources
 Rice Genome Research Program
 2-1-2 Kannondai, Tsukuba
 Ibaraki,
 Japan 305
 Tel: 0298-38-7441
 Fax: 0298-38-7468
 Email: tsasaki@abr.affrc.go.jp
 PROJECT = 'RGP'

FEATURES
 source Location/Qualifiers
 1..567
 /organism="Oryza sativa subsp. japonica"
 /strain="cultivar Nipponbare, sub-species japonica"
 /db_xref="taxon:39947"
 /clone="C0777.102"
 /note="lib="Rice callus"
 /note="Vector: pBluescript II SK+; Site_1: SalI; Site_2:
 NotI; cDNA prepared from rice callus mRNAs by using
 oligo(dT) as a primer and ligating to the SalI-NotI site
 of pBluescript II SK+ phagemid." 1 others

BASE COUNT 190 a 105 c 196 g 75 t
 ORIGIN

Query Match 60.0%; Score 15; DB 156; Length 567;
 Best Local Similarity 100.0%; Pred. No. 73;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 tgcacgcacgacatt 20
 |||||||||||||
 Db 472 TGCATGCCGACGATT 458

RESULT 44
 AA405993/c 581 bp mRNA EST 09-NOV-1997
 LOCUS Z066f09.r1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:742985
 DEFINITION 5' similar to contains MER29.b3 MER29 repetitive element ;, mRNA
 sequence.
 ACCESSION AA405993
 VERSION AA405993.1 GI:2063976
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 581)
 AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisels,G., Jost,S.,
 Kitzman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin
 White,Y., Wylie,T., Waterston,R. and Wilson,R.
 JOURNAL WASHU-MCI human EST Project
 COMMENT Unpublished (1997)
 Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert Length: 997 Std Error: 0.00
 Seq primer: -28m13 rev2 ET from Amersham
 High quality sequence stop: 498.

FEATURES
 source Location/Qualifiers
 1..581
 /organism="Homo sapiens"
 /db_xref="GDB:5930030"
 /db_xref="taxon:9606"
 /clone="IMAGE:742985"
 /clone_lib="Soares_testis_NHT"
 /sex="male"
 /lab_host="DH10B"
 /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
 was prepared from mRNA obtained from Clontech laboratories
 , Inc., and primed with a Not I - oligo(dT) primer [5'
 TGTTCACATCTGACAGTGGACGGCCGCCCAATTCTTTTCTTTT 3']
 Double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pT7T3 vector. Library
 went through one round of normalization to Cot5, and was
 constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 153 a 171 c 131 g 126 t
 ORIGIN

Query Match 60.0%; Score 15; DB 7; Length 581;
 Best Local Similarity 100.0%; Pred. No. 73;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 catcgccacgatttc 22
 |||||||||||||
 Db 81 CATCGCCGACGATTTC 67

RESULT 45
 A0783596 605 bp DNA GSS 03-AUG-1999
 LOCUS HS_3121_A2_F07_T7A CIT Approved Human Genomic Sperm Library D Homo
 DEFINITION sapiens genomic clone Plate=3121 Col=14 Row=K, DNA sequence.
 ACCESSION A0783596
 VERSION A0783596.1 GI:5691150
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 605)
 AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
 Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
 Hood,L.
 Sequence-tagged connectors: A sequence approach to mapping and
 scanning the human genome
 Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 JOURNAL 99380589
 MEDLINE
 COMMENT Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallaceu.washington.edu
 Clones may be purchased from Research Genetics (info@resgen.com).
 BAC end Web Server: http://www.htsc.washington.edu
 Plate: 3121 row: K column: 14
 Seq primer: T7
 Class: BAC ends
 High quality sequence stop: 605.

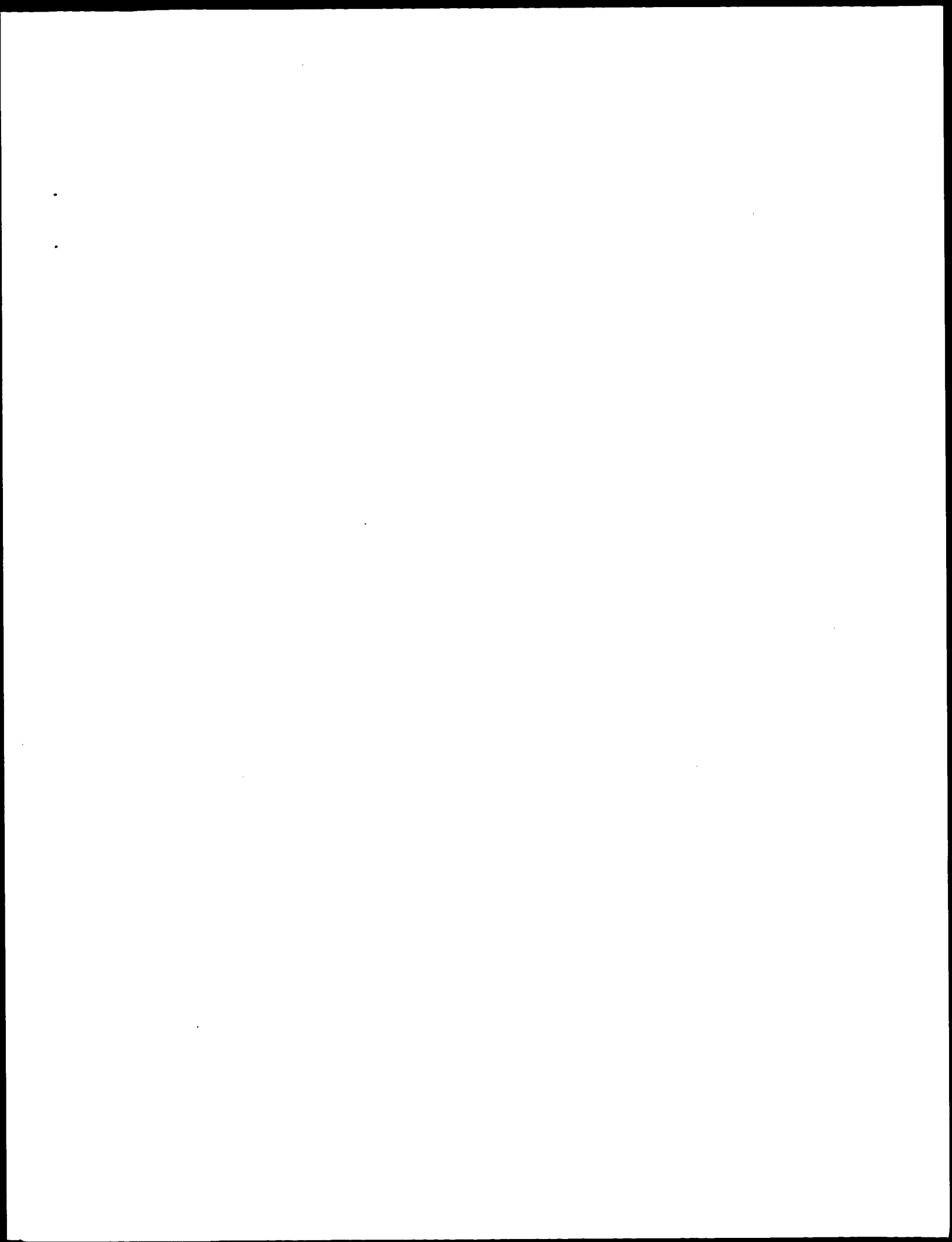
FEATURES
 source Location/Qualifiers
 1..605
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="Plate=3121 Col=14 Row=K"
 /clone_lib="CIT Approved Human Genomic Sperm Library D"
 /sex="male"
 /note="Organ: sperm; Vector: pBelobAC11; BAC Clones in

BASE COUNT E-Coli DH10B"
ORIGIN 187 a 187 c 123 g 92 t 16 others

Query Match 60.0%; Score 15; DB 233; Length 605;
Best Local Similarly 100.0%; Pred. No. 73;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 gtgcatgccagcat 19
 |||||
Db 248 GTGCATGCCAGCAT 234

Search completed: October 9, 2001, 15:15:32
Job time: 13654 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 15:46:36 ; Search time 1670.83 seconds
(without alignments)
231.438 Million cell updates/sec

Title: US-09-396-196F-7

Perfect score: 25
Sequence: 1 gcatcgccagcattcgatctcgt 25

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1344157 seqs, 7733874588 residues

Word size : 9

Total number of hits satisfying chosen parameters: 201657

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl: *
1: gb_ba1: *
2: gb_ba2: *
3: gb_ba3: *
4: gb_in1: *
5: gb_in2: *
6: gb_in3: *
7: gb_om: *
8: gb_ov: *
9: gb_pat1: *
10: gb_pat2: *
11: gb_ph: *
12: gb_pl1: *
13: gb_pl2: *
14: gb_pl3: *
15: gb_pl4: *
16: em_ba1: *
17: em_ba2: *
18: em_fun: *
19: em_htgo_hum: *
20: em_htgo_inv: *
21: em_htgo_rod: *
22: em_htg_hum1: *
23: em_htg_hum2: *
24: em_htg_hum3: *
25: em_htg_hum4: *
26: em_htg_hum5: *
27: em_htg_hum6: *
28: em_htg_hum7: *
29: em_htg_hum8: *
30: em_htg_inv1: *
31: em_htg_inv2: *
32: em_htg_other: *
33: em_htg_rod: *
34: em_hum1: *
35: em_hum2: *
36: em_hum3: *
37: em_hum4: *
38: em_hum5: *
39: em_hum6: *
40: em_hum7: *
41: em_in: *
42: em_om: *
43: em_or: *

44: em_ov: *
45: em_pat: *
46: em_ph: *
47: em_pl: *
48: em_ro: *
49: em_sts: *
50: em_sy: *
51: em_un: *
52: em_vl: *
53: gb_sts1: *
54: gb_sts2: *
55: gb_sts3: *
56: gb_sy: *
57: gb_un: *
58: gb_vl1: *
59: gb_vl2: *
60: gb_htg1: *
61: gb_htg2: *
62: gb_htg3: *
63: gb_htg4: *
64: gb_htg5: *
65: gb_htg6: *
66: gb_htg7: *
67: gb_htg8: *
68: gb_htg9: *
69: gb_htg10: *
70: gb_htg11: *
71: gb_htg12: *
72: gb_htg13: *
73: gb_htg14: *
74: gb_htg15: *
75: gb_htg16: *
76: gb_htg17: *
77: gb_htg18: *
78: gb_htg19: *
79: gb_htg20: *
80: gb_htg21: *
81: gb_htg22: *
82: gb_htg23: *
83: gb_htg24: *
84: gb_htg25: *
85: gb_pr1: *
86: gb_pr2: *
87: gb_pr3: *
88: gb_pr4: *
89: gb_pr5: *
90: gb_pr6: *
91: gb_pr7: *
92: gb_pr8: *
93: gb_pr9: *
94: gb_r01: *
95: gb_r02: *
96: gb_in4: *
97: gb_pr10: *
98: em_ba3: *

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	25	100.0	1041	9	AR029499	AR029499 Sequence
2	25	100.0	1041	9	AR034916	AR034916 Sequence
3	25	100.0	1084	9	A11530	A11530 BiOB gene o
4	25	100.0	1121	10	E00893	E00893 Genomic DNA
5	25	100.0	5793	2	ECOBIO	J04423 E.coli 7-8-
6	25	100.0	5872	9	A38246	A38246 Sequence 1
7	25	100.0	5872	9	A38251	A38251 Sequence 6
8	25	100.0	5872	9	A93674	A93674 Sequence 1

```

9      25      100.0      5872      9      A93679      Sequence 6
10     25      100.0      5872      9      AR101809      Sequence
11     25      100.0      5872      9      AR101810      Sequence
12     25      100.0      11022     1      AE000180      Sequence
13     25      100.0      11501     1      AE005258      Escherich
14     25      100.0      297816    2      AP002553      Escherich
15     24      96.0       5526     2      AF250776      Escherich
16     18      72.0       128      2      EC0108      Unculture
17     17      68.0       479     54      G0146      Escherichia
18     17      68.0      53067     65      AC019940      Dros
19     17      68.0      153870    63      AC013451      Drosophill
20     17      68.0      183439     4      AC007809      Drosophill
21     17      68.0      232744     5      AE003705      Drosophill
22     17      68.0      237119     65      AC017740      Drosophill
23     17      68.0      327446     4      AE003500      Drosophill
24     16      64.0      1186     2      AF146614      Ertvinia c
25     16      64.0      11910     56      AF206717      Shutle v
26     16      64.0      117968     92      HS272E8      293929 Human DNA s
27     16      64.0      118100     65      AC017911      Drosophill
28     16      64.0      140825     13      AP002817      Oryza sat
29     16      64.0      146081     13      AP001366      Oryza sat
30     16      64.0      171705     4      AC007476      Drosophill
31     16      64.0      183937     71      AC037435      Mus muscu
32     16      64.0      266133     5      AE003822      Drosophill
33     15      60.0      1389     14      NTPTLG      X61102 N. tabacum q
34     15      60.0      1713     14      NTPECL1      X61158 N. tabacum g
35     15      60.0      1907     93      HSSYNTRO5      AJ003029 Homo sapi
36     15      60.0      2616     85      AB048915      Macaca fa
37     15      60.0      3677     3      PSTAAMH      M11035 P. syringae
38     15      60.0      5618     85      AB028980      Homo sapi
39     15      60.0      10205     1      AE004320      AE004320 Vibrio ch
40     15      60.0      10593     1      AE003860      Xylella f
41     15      60.0      10978     1      AE004795      Pseudomon
42     15      60.0      23873     65      AC018333      Drosophill
43     15      60.0      31424     58      AF168792      Gallid he
44     15      60.0      57237     6      CEY6E10A      A113285 Caenorhab
45     15      60.0      94487     12      AC012394      Arabidops

```

ALIGNMENTS

```

RESULT 1
LOCUS   AR029499      1041 bp      DNA
DEFINITION   Sequence 7 from patent US 5859335.
ACCESSION   AR029499
VERSION     AR029499.1      GI:5941472
KEYWORDS
SOURCE      Unknown.
ORGANISM    Unclassified.
REFERENCE   1 (bases 1 to 1041)
AUTHORS    Patton,D.Andrew.
TITLE      Enhanced biotin biosynthesis in plant tissue
JOURNAL    Patent: US 5859335-A 7 12-JAN-1999;
FEATURES
     source             1..1041
                        /organism="unknown"
BASE COUNT      262 a      273 c      305 g      201 t
ORIGIN
Query Match      100.0%; Score 25; DB 9; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.00022;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1      gcatcgccagcattcgatcctcgt 25
      ||||||||||||||||||||||||||||
Db      90      GCATCGCCAGCATTGATCCTCGT 114
RESULT 2      '

```

```

AR034916
LOCUS   AR034916      1041 bp      DNA
DEFINITION   Sequence 7 from patent US 5869719.
ACCESSION   AR034916
VERSION     AR034916.1      GI:5950521
KEYWORDS
SOURCE      Unknown.
ORGANISM    Unclassified.
REFERENCE   1 (bases 1 to 1041)
AUTHORS    Patton,D.A.
TITLE      Transgenic plants having increased biotin content
JOURNAL    Patent: US 5869719-A 7 09-FEB-1999;
FEATURES
     source             1..1041
                        /organism="unknown"
BASE COUNT      262 a      273 c      305 g      201 t
ORIGIN

```

```

Query Match      100.0%; Score 25; DB 9; Length 1041;
Best Local Similarity 100.0%; Pred. No. 0.00022;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1      gcatcgccagcattcgatcctcgt 25
      ||||||||||||||||||||||||||||
Db      90      GCATCGCCAGCATTGATCCTCGT 114

```

```

RESULT 3
LOCUS   A11530      1084 bp      DNA
DEFINITION   Bior gene of E.coli with primers.
ACCESSION   A11530
VERSION     A11530.1      GI:490218
KEYWORDS
SOURCE      Escherichia coli.
ORGANISM    Escherichia coli.
REFERENCE   1 (bases 1 to 1084)
AUTHORS
JOURNAL    Patent: GB 2216530-A 16 11-OCT-1989;
FEATURES
     source             1..1084
                        Location/Qualifiers
                        1..1084
                        /organism="Escherichia coli"
                        /db_xref="taxon:562"
                        24..1064
                        /gene="BIOB"
                        24..1064
                        /gene="BIOB"
                        /codon_start=1
                        /transl_table=1
                        /protein_id="CA00965.1"
                        /db_xref="GI:490219"
                        /db_xref="GI:490219"
                        /translation="MAHRPMTLSOYELFEKPLLDLFEAQOYHROHPDROYAST
                        LLSITGACPEDCYKCPROSSRYKGLAEERLMEYQLBESARRAKAGSTRPGGAAM
                        KNHPRDMPYIEOWVQGVKAMGEACHTLTGLSSQAKORLANAGLDYINHLDTSPF
                        YGNITTRTYQERLDLTLEKADAGIKVCSGIVGLGTVKDRAGLLQLANLPTPPS
                        VPINMLVKVGTPLADNDVDVAFIFRTIIVARIIMPTSYVRLSAGREOMNEDTOAMC
                        EMAGANSIFYGCKLLTTPNPEEDKDLQLFKRLGINPQTFVLAVGDNQQRRLDQALMT
                        PDTDEYVNAAL"
BASE COUNT      271 a      286 c      318 g      209 t
ORIGIN

```

```

Query Match      100.0%; Score 25; DB 9; Length 1084;
Best Local Similarity 100.0%; Pred. No. 0.00022;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1      gcatcgccagcattcgatcctcgt 25
      ||||||||||||||||||||||||||||

```

Db 113 GCATGCCAGCATTTCGATCTCGT 137

RESULT 4

LOCUS E00893 1121 bp DNA PAT 29-SEP-1997

DEFINITION Genomic DNA encoding biotin Synthetase.

ACCESSION E00893

VERSION E00893.1 GI:2169154

KEYWORDS JP 1986149091-A/1.

SOURCE Escherichia coli.

ORGANISM Escherichia coli.

Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.

REFERENCE 1 (bases 1 to 1121)

AUTHORS Hirono, Y., Kojima, T. and Kimura, H.

TITLE DUPLEX DNA TO CODE BIOTIN SYNTHASE, BACTERIUM CONTAINING SAME AND PRODUCTION OF BIOTIN

JOURNAL Patent: JP 1986149091-A 1 07-JUL-1986;

COMMENT NIPON SODA CO LTD

OS Escherichia coli

PN JP 1986149091-A/1

PD 07-JUL-1986

PF 24-DEC-1984 JP 1984272605

PI HIRONO YOSHIIKO, KOJIMA TAKAKAZU, KIMURA HITOSHI PC

C12N15/00,C12N1/20,C12P13/18,(C12N1/20,C12P13/18, PC

C12R1:19);

CC strandedness: Double;

CC topology: Linear;

CC hypothetical: No;

CC anti-sense: No;

CC *Source: Strain-Escherichia coli Nsl01;

CC Feature is identified by experimental;

CC Key Location/Qualifiers

FT CDS 42..1079

FT /product="biotin synthetase".

FEATURES

source 1..1121

/organism="Escherichia coli"

/db_xref="taxon:562"

BASE COUNT 289 a 296 c 325 g 211 t

ORIGIN

Query Match 100.0%; Score 25; DB 10; Length 1121;

Best Local Similarity 100.0%; Pred. No. 0.00022;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gcatgccagcatttcgattcgcgt 25

|||||

Db 131 GCATGCCAGCATTTCGATCTCGT 155

|||||

RESULT 5

LOCUS E00893 5793 bp DNA BCT 28-FEB-1994

DEFINITION E.coli 7,8-diamino-pelargonic acid (bioA), biotin synthetase (bioB), 7-keto-8-amino-pelargonic acid synthetase (bioF), bioc protein, and dehydrobiotin synthetase (bioD), complete cds.

ACCESSION J04423

VERSION J04423.1 GI:145422

KEYWORDS 7,8-diamino-pelargonic acid aminotransferase; 7-keto-8-amino-pelargonic acid synthetase; bioA gene; bioB gene; bioc gene; bioD gene; bioF gene; biotin synthetase; dehydrobiotin synthetase;

SOURCE Escherichia coli (strain K-12) DNA.

ORGANISM Escherichia coli

Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.

REFERENCE 1 (bases 1 to 5793)

AUTHORS Otsuka, A., U., Buoncristiani, M.R., Howard, P.K., Flamm, J. and Johnson, O.

TITLE The Escherichia coli biotin biosynthetic enzyme sequences

JOURNAL U. Biol. Chem. 263, 19577-19585 (1988)

MEDLINE 89066784

COMMENT Draft entry and computer-readable sequence [1] kindly submitted by A.Otsuka, 09-NOV-1988.

FEATURES

source Location/Qualifiers

1..5793

/organism="Escherichia coli"

/strain="K-12"

/db_xref="taxon:562"

/complement(98..574)

/note="ORF 1"

/codon_start=1

/transl_table=1

/protein_id="AAA23513.1"

/db_xref="GI:45168"

/translation="MKLISNDLRDGLPHRHVNGMGVGDGDNISPLHAWDVPAGTK SEVVCYDPDAPGSGMMHVVVNLPAOTRYLPQGFSGLVAMPDGYLOTPTDGGKTG YDGAAPRGETHRYIFTVHALDIERIDVDGASAMVGFNVHFSLASAITAMFS"

/complement(633..1925)

/gene="bioA"

/complement(633..1925)

/gene="bioA"

/codon_start=1

/transl_table=1

/product="7,8-diamino-pelargonic acid aminotransferase"

/protein_id="AAA23514.1"

/db_xref="GI:457106"

/translation="MTTDLAPDQRHIMHPTYSMTSPLPVYPVSAECCILSDGR LVDGSMWMAAIIHGYNHPQLNNAKMSQIDAMSHVWGGITNAPALELCKLVAMSGRN ALCEVLADSGSVAVEAVAMKALQYWKQGEARQPTLFERNYHGDGFGASVCDPN SMHSLMKGYLPNLFPAPQSRMDGEMDRDVGARLAARHIAHVEIAVIEPTVOGA GCMRHWPMPLRIKRICDRGILLIADIEINFGPRTGCFACAEAASTADILCSGKA LTGGTWTLSATITTRVARTISNGRAGCPMHPGPMPLCAANAATLIESGDMO QGVADLEVQLRLAPARDAEAVADVRLGATGVETTHPVMAALQKFPVQGVWIR PGKLIYIMPYIILPQQLRLAAVNAVVDTEFCQ"

2012..3052

/gene="bioB"

2012..3052

/gene="bioB"

/codon_start=1

/transl_table=1

/product="biotin synthetase"

/protein_id="AAA23515.1"

/db_xref="GI:145425"

/translation="MAHRRRWLTISOYTELEFEKPLDILFEAOQVHRHPPROQVST LLSIKTGACPEDECKTPQTSRYKTGIEARILMEVEYVLESARKARAAGSTRCCMAAM KNPHERDMPEYLDQWGVGRAMGLEACMTLGTLSQAOQLANAGLDIYNHNDTPER YGNIITRTYQERLDLTLEKVRDAGIKVCSGGIVGLGTEYKVRAGLLDLQLANPTPES VPINMLVKVGFPLADNDVDADFIRITAVARIIMPTSVYRLSAGREOMDQOAMC PMAGANSIFYGCKLFTTPNDEBDKDLQRLFRKIGLNPQYRAVLADNEOQORLEQALMT PPDDEYYNAAL"

3049..4203

/gene="bioF"

3049..4203

/gene="bioF"

/codon_start=1

/transl_table=1

/product="7-keto-8-amino pelargonic acid synthetase"

/protein_id="AAA23516.1"

/db_xref="GI:145426"

/translation="MSMOEKINAAIDARRAADAARRRYPAVGAGRWLVADROLYINF SSNDYGLSHNPQIIRAMOGAOGFISGSGSHVGSYVAHQLEEELEAEWLGYSRA LPTISGPAANOAVIYAAMKEDRIADRLSHASLLPAASISPSQLRFRPHNDVTHLAR LLSAPCGQGOVYVTEGVESMDGSAPLAETQOYTOOHNGMIAWDHAGHVGISEGGRG SCHLOKVRPELVVYTGKRGVSGAIVLCSSTYADYLLQFARHLIYSTMPAPAOAL RASLAVIRSDGARRREKLAALITFRAGVQDLPFTLABSCSAIOPLYICDNRALQI AEKLRQGCWVTAIRPTVPAGTARLRLTLTAHEMODIDRLLEVLIHNG"

4190..4945

/gene="bioc"

4190..4945

/gene="bioc"

/gene="bioc"


```

/evidence=experimental
/product="PROTEIN"
/protein_id="CAA02328.1"
/db_xref="GI:2294848"
/translation="MKLISNDLRDGDKILPHRVFNGMGVGDGNISPHLAWDDVPAGTK
STVVTCTYDDAPFTGSGMHWVYVNLPAOTRVLPQGFSGLVAMPDGVQOTRDPCKTG
YDGAAPPKGETHRVIFVYHALDIERIDVEGASGAMVGFNVHFSLASASTAMPS"
stem_loop
5583..5605
5583..5644
terminator
/standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
TERMINATOR"
BASE COUNT      1318 a      1552 c      1695 g      1307 t
ORIGIN

Query Match      100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.0002;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 gcatcgccagcattcgatccgcgt 25
        |||||||||||||||||||
Db      206 GCATCGCCAGCATTTGCATCTCGT 230

RESULT 7
LOCUS      A38251      5872 bp      DNA      PAT      05-MAR-1997
DEFINITION      Sequence 6 from Patent WO9408023.
ACCESSION      A38251
VERSION      A38251.1 GI:2294849
KEYWORDS
SOURCE      Escherichia coli.
ORGANISM      Escherichia coli.
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE      1 (bases 1 to 5872)
AUTHORS      Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
TITLE      BIOTECHNOLOGICAL METHOD OF PRODUCING BIOTIN
JOURNAL      Patent: WO 9408023-A 6 14-APR-1994;
            LONZA AG (CH)
COMMENT      Other publication PL 308301 950724
            Other publication CA 2145400 940414
            Other publication AU 4820293 940426
            Other publication HU 71781 960228
            Other publication SK 42095 951108
            Other publication CZ 9500809 950913
            Other publication FI 951547 950331
            Other publication JP 8501694* 960227.
            location/Qualifiers
FEATURES
    source
        1..5872
            /organism="Escherichia coli"
            /strain="DSM498"
            /db_xref="taxon:562"
            /clone="PBO30A15-9"
            1141..1156
            /standard_name="BIOF RBS"
            1154..2308
            /gene="BIOF"
            1154..2308
            /gene="BIOF"
            /standard_name="8-AMINO-7-OXONONANOATE SYNTHASE"
            /EC_number="2.3.1.47"
            /codon_start=1
            /transl_table=11
            /number=2
            /evidence=experimental
            /product="KAPA SYNTHASE"
            /protein_id="CAA02329.1"
            /db_xref="GI:2294850"
            /translation="MSMOEKINALLDARRAADALRRRYPAAGAGRMLVADDDRYLNF
            SSNDYLGISHHPDIIIRAMQOGAEQFGISGGSGVSGSVYHQALEELBWLSTSA
            LLPTISGFANQAVIAAMAKEDRIIAURLSHASLLEASLSPQLRRFAHNDVYHLAR
            LLASPCGQGMVTEGVFSGMDGSAPLAETIOOVYQOHNGWIMVDAGTGVIGBGRG
            SCMLQKVPKELLVYTFGKGFVSGAIVICSTYADYLLQFARHLITYSMPAPAOAL
            RASIAVIRSDGDKRRREKIALILTRFPAGVODI.PEPLIADSCSAIOPLIVDNRALQL
            AEKLRQOGCVTAIRPPTVPAGTARLMLTLAAHEMODIDRLLEVLHGNG"
            3030..3045
            /standard_name="BIOD RBS"
            3043..3753
            /gene="BIOB"
            3043..3753
            /gene="BIOB"
            /standard_name="DETHIOBIOTIN SYNTHASE"
            /EC_number="6.3.3.3"
            /codon_start=1
            /transl_table=11
            /number=4
            /evidence=experimental
            /product="PTB SYNTHASE"
            /protein_id="CAA02330.1"
            /db_xref="GI:2294851"
            /translation="MSKRYEVTGDTVEGKTVAACALQAAKAGRYTAGYKPYASGS
            EKTPEGLRNSDALIQRNSSIQLDYATVNPYTPAEPSPHISAEGRPIESLYMSAG
            LRALFQADWYLVBEAGGMPPLSDPTFPAWYQEOPLVILVYVIGKGINHMLTA
            QVIOHAGLTLGWAVANDVTPPGKRAEYMTTLTNITAPILGELPMLAENENATGK
            YINAEVDSYLGFTSRL"
BASE COUNT      1318 a      1552 c      1695 g      1307 t
ORIGIN

Query Match      100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.0002;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 gcatcgccagcattcgatccgcgt 25
        |||||||||||||||||||
Db      206 GCATCGCCAGCATTTGCATCTCGT 230

RESULT 8
LOCUS      A93674      5872 bp      DNA      PAT      22-JAN-2000
DEFINITION      Sequence 1 from Patent EP0798384.
ACCESSION      A93674
VERSION      A93674.1 GI:6741862
KEYWORDS
SOURCE      Escherichia coli.
ORGANISM      Escherichia coli.
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE      1 (bases 1 to 5872)
AUTHORS      Birch,O. and Brass,J.
TITLE      Biotechnological method of producing biotin
JOURNAL      Patent: EP 0798384-A 1 01-OCT-1997;
            LONZA AG (CH)
COMMENT      location/Qualifiers
FEATURES
    source
        1..5872
            /organism="Escherichia coli"
            /strain="DSM498"
            /db_xref="taxon:562"
            /clone="PBO30A-15/9"
            1..96
            /function="PROMOTER PTAC"
            /evidence=experimental
            23..28
            /evidence=experimental
            45..49
            /standard_name="PROMOTER PTAC"
            /evidence=experimental
            105..119
            /standard_name="BIOB RBS NO.9"
            /evidence=experimental
            117..1157
            /gene="BIOB"
            117..1157
            /gene="BIOB"
            /codon_start=1
    promoter
        -35_signal
        -10_signal
    RBS
        105..119
            /standard_name="BIOB RBS NO.9"
            /evidence=experimental
            117..1157
            /gene="BIOB"
            117..1157
            /gene="BIOB"
            /codon_start=1
    gene
        117..1157
            /gene="BIOB"
            /codon_start=1
    CDS
        117..1157
            /gene="BIOB"
            /codon_start=1

```

```

/transl_table=11
/evidence=experimental
/product="BIOTIN SYNTHASE"
/protein_id="CAB69590.1"
/db_xref="GI:6741863"
/translation="MAHRPMTLSOVTELEFPEKPLDLLFEAQOVHROHDPVOVST
LSTKTCACPEDCKYCPOSSRYKTGDEAEKLEVEVLESARAKAGSTRPCMGAW
KNFERDMPYLEOVGVKAMGLEACWTGTLSESOAQLANGLDYNNLDTSPER
YGNITRTYQERLDLEKVDAGIKVSGGIVGLGPTVAGLLOLANLEPTPES
VPIMLVKVGTPLADNDVDADFIRITIAVARIMPTSVYRSGAGEOWEOTAMC
FMAGANSIFYGCKLITTPNEEDKDLQLFKRLGILNPQGTAVLAGDNBQOORLEQALMT
PDIDEYNAAL"
2284..2297
/standard_name="BIOC RBS"
2295..3050
/gene="BIOC"
2295..3050
/gene="BIOC"
/function="INVOLVED IN PIMELOYL-COA SYNTHESIS"
/codon_start=1
/transl_table=11
/product="PROTEIN"
/protein_id="CAB69591.1"
/db_xref="GI:6741864"
/translation="MATVVKQALTAARGRAAHYBOHADLQOSADALLAMPORKYT
HVIDAGCGPGMSRHWRRRAQYTAIDISPPMLVOROKDAADHYLAGDIESIPLATA
TTPDLSANLAVOMCGNLSALRELAVVPRKGVAVFTLVGSLAPELHQAQWADERP
HANRELPPDEIFOSLVGHVQHIIQPIITTFEDALSMSRLKGIGATHLHEGRDRIIL
TRSOQLQLOLAMPQOQGRYPPLYHFLGVIARE"
3742..3752
/standard_name="BIOA RBS"
3750..5039
/gene="BIOA"
3750..5039
/gene="BIOA"
/standard_name="S-ADENOSYL-L-METHIONINE:8-AMINO-7-
OXONANONATE AMINOTRANSF."
/ec_number="2.6.1.62"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="DAPA SYNTHASE"
/protein_id="CAB69592.1"
/db_xref="GI:6741865"
/translation="MTTDLADFDRHIHMYPTSMTPLPVYPVASBEGCELLSDGR
LYDGMSSMAAIIHGYNNHLPOLNAKMSODAMSHVMGGITTHAPATELCRKIYAMROP
LECVLADGSVAVEAVEMALQYMOAGKAEORRLTRNGHGTPTGMSVCDPDS
MSIMKGYLPENILFAPAROSRMDGEMDERDMYGRMLMAHREHETAAVITEIVGAG
GKRYHPEWLKRIKICDEGILIADEIATGFRGKILFACENAEIADILCLKAL
TGGIMTLSATLTREVAETISNGEGCEFMHGPTFMGNPLACAAANASLAILESSGMQO
QVADIEVOLREQLARDAEMVADVIVGALGVETTHPVNMAALOKFVEGVWIRP
FGKLIYMPPIIILPOOLRLTAANRAVQDETFCQ"
5088..5100
/standard_name="ORFI RBS"
5098..5574
/gene="ORFI"
5098..5574
/gene="ORFI"
/function="UNKNOWN, INVOLVED IN BIOTIN SYNTHESIS"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="PROTEIN"
/protein_id="CAB69593.1"
/db_xref="GI:6741866"
/translation="MKLISNDLRDGLPHRRHVFNGMGYDGDNISPHLAMDVAGTK
SEVVTCYDPAPTGSGMHWVYVNLPAETRVLPQSGSLVAMPDVILQITTTGKTS
YDGAAPKRGEMHRYIFTHALDIERIVDEBAGSAMGCFNVHFSLASASITAMFS"
5583..5605
/standard_name="RHO-INDEPENDENT TRANSCRIPTIONAL
TERMINATOR"
1318 a, 1552 c 1695 g 1307 t

```

```

ORIGIN
Query Match          100.0%; Score 25; DB 9; Length 5872;
Best Local Similarity 100.0%; Pred. No. 0.0002;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1  gcatgccagattcgatccgt 25
Db      206 GCATGCCAGACATTTCATCTCGT 230

RESULT      9
A93679      A93679      5872 bp      DNA      PAT      22-JAN-2000
LOCUS      Sequence 6 from Patent EP0798384.
DEFINITION  A93679
ACCESSION  A93679
VERSION    A93679.1  GI:6741867
KEYWORDS   .
SOURCE     Escherichia coli.
           Escherichia coli
           Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
           Escherichia.
REFERENCE   1 (bases 1 to 5872)
            Birch, O. and Brass, J.
            Biotechnological method of producing biotin
            Patent: EP 0798384-A 6 01-OCT-1997;
            LONZA AG (CH)

FEATURES
     source          location/Qualifiers
     RBS             1..5872  /organism="Escherichia coli"
     gene            /strain="DSM498"
     CDS             /db_xref="taxon:562"
                     /clone="PB030A15-9"
                     1141..1156
                     /standard_name="BIOF RBS"
                     1154..2308
                     /gene="BIOF"
                     1154..2308
                     /gene="BIOF"
                     /standard_name="8-AMINO-7-OXONANONATE SYNTHASE"
                     /ec_number="2.3.1.47"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="KAPA SYNTHASE"
                     /protein_id="CAB69594.1"
                     /db_xref="GI:6741868"
                     /translation="MSWQEKINALLDARRAADALRRRYPAAGAGRMIVADPOYINF
                     SSNDYGLSHHPQIIRAMQOAGOPFGISGSGSHVGYVHQALEELAEMLGYSRA
                     LLPTSGRAANDAVIYAAMAKEDITADRLSHASLLRASISPSQLRFRANDVTHLAR
                     LLASPCGQOVVTEGVFSMDGDSAPLAIQOVYTOQNGMLWDDHAGVGEQDGRG
                     SCHLQKVPPELVVTFEGFGVSGAVALSSVYADVLDFARHLIYSTPMPAQAQAL
                     RASLAVTRSDGDAKREKLAALLTRFRAGVODLPFTIADSCSAIOPLIYDMSRALOL
                     AEKIROGCGWVTAIRPTVPAGTARLRLLTAAHEMODIDRLLEVLHGNG"
                     3030..3045
                     /standard_name="BIOB RBS"
                     3043..3753
                     /gene="BIOB"
                     3043..3753
                     /gene="BIOB"
                     /standard_name="DETHIOBIOTIN SYNTHASE"
                     /ec_number="6.3.3.3"
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="DTB SYNTHASE"
                     /protein_id="CAB69595.1"
                     /db_xref="GI:6741869"
                     /translation="MSKRYFVGTQTEVGVASCALLQAAKAGFTAGYRPAVSGS
                     EKPEGLRNSDALALORNSSLQIDATVAVPTTFAETSPHITSADGRITESLVASAG
                     LRLBOQADWVLVEGAGWFTPLSDFTFADWTOEQLVILVGVKLCGINHAMLTA
                     OVTLRHAGLTLAGVANDVTPPKRRRAEYMTLLTRMIPALDLGEIPLAENPNNAATGK

```

YINLAFVDASTLCFTSRL"
 BASE COUNT 1318 a 1552 c 1695 g 1307 t
 ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 5872;
 Best Local Similarity 100.0%; Pred. No. 0.0002;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcatgccagcattcgatccctg 25
 |||||||||||||||||||||||||
 Db 206 GCATGCCAGCATTGCGATCCTCGT 230

RESULT 10
 ARI01809
 LOCUS ARI01809 5872 bp DNA PAT 14-FEB-2001
 DEFINITION Sequence 1 from patent US 6083712.
 ACCESSION ARI01809
 VERSION ARI01809.1 GI:12812607
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 5872)
 AUTHORS Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
 TITLE Biotechnological method of producing biotin
 JOURNAL Patent: US 6083712-A 1 04-JUL-2000;
 FEATURES Location/Qualifiers
 source 1..5872

BASE COUNT 1318 a 1552 c 1695 g 1307 t
 ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 5872;
 Best Local Similarity 100.0%; Pred. No. 0.0002;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcatgccagcattcgatccctg 25
 |||||||||||||||||||||||||
 Db 206 GCATGCCAGCATTGCGATCCTCGT 230

RESULT 11
 ARI01810
 LOCUS ARI01810 5872 bp DNA PAT 14-FEB-2001
 DEFINITION Sequence 6 from patent US 6083712.
 ACCESSION ARI01810
 VERSION ARI01810.1 GI:12812608
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unclassified.
 REFERENCE 1 (bases 1 to 5872)
 AUTHORS Birch,O., Brass,J., Fuhrmann,M. and Shaw,N.
 TITLE Biotechnological method of producing biotin
 JOURNAL Patent: US 6083712-A 6 04-JUL-2000;
 FEATURES Location/Qualifiers
 source 1..5872
 BASE COUNT 1318 a 1552 c 1695 g 1307 t
 ORIGIN

Query Match 100.0%; Score 25; DB 9; Length 5872;
 Best Local Similarity 100.0%; Pred. No. 0.0002;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcatgccagcattcgatccctg 25
 |||||||||||||||||||||||||
 Db 206 GCATGCCAGCATTGCGATCCTCGT 230

RESULT 12
 AE000180

LOCUS AE000180 11022 bp DNA BCT 01-DEC-2000
 DEFINITION Escherichia coli K12 MG1655 section 70 of 400 of the complete
 genome.
 ACCESSION AE000180 000096
 VERSION AE000180.1 GI:1786988
 KEYWORDS
 SOURCE Escherichia coli K12.
 ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Escherichia.

REFERENCE 1 (bases 1 to 11022)
 AUTHORS Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V.,
 Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F.,
 Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J.,
 Mau,B. and Shao,Y.
 TITLE The complete genome sequence of Escherichia coli K-12
 JOURNAL Science 277 (5331), 1453-1474 (1997)
 MEDLINE 97426617
 PUBMED 9278503

REFERENCE 2 (bases 1 to 11022)
 AUTHORS Blattner,F.R.
 TITLE Direct Submission
 JOURNAL Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
 University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
 Email: ecollegenetics.wisc.edu Phone: 608-262-2534 Fax:
 608-263-7459

REFERENCE 3 (bases 1 to 11022)
 AUTHORS Blattner,F.R.
 TITLE Direct Submission
 JOURNAL Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
 University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
 Email: ecollegenetics.wisc.edu Phone: 608-262-2534 Fax:
 608-263-7459

REFERENCE 4 (bases 1 to 11022)
 AUTHORS Plunkett,G. III.
 TITLE Direct Submission
 JOURNAL Submitted (13-OCT-1998) Laboratory of Genetics, University of
 Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
 This sequence was determined by the E. coli Genome Project at the
 University of Wisconsin-Madison (Frederick R. Blattner, director).
 Supported by NIH grants HG00301 and HG01428 (from the Human Genome
 Project and NCHGR). The entire sequence was independently
 determined from E. coli K12 strain MG1655. Predicted open reading
 frames were determined using Genemark software, kindly supplied by
 Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
 30332 (e-mail: markember.gatech.edu). Open reading frames that
 have been correlated with genetic loci are being annotated with CG
 Site Nos., unique ID nos. for the genes. In the E. coli Genetic
 Stock Center (CGSC) database at Yale University, kindly supplied by
 Mary Berlyn. A public version of the database is accessible
 (http://cgsc.biology.yale.edu). Annotation of the genome is an
 ongoing task whose goal is to make the genome sequence more useful
 by correlating it with other data. Comments to the authors are
 appreciated. Updated information will be available at the E. coli
 Genome Project's World Wide Web site
 (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and
 its annotations are periodically updated; this is version M54. No
 sequence changes. Annotation updates: updated gene identifications
 and products; all new functional assignments courtesy of Monica
 Riley; added promoters, protein binding sites, and repeated
 sequences described in reference 1. The unique numeric identifiers
 beginning with a lowercase 'b' assigned to each gene (protein- or
 RNA-encoding) are now designated as gene synonyms instead of
 labels. This should allow them to be searched for in Entrez as gene
 names.

FEATURES
 source

Location/Qualifiers
 1..11022
 /organism="Escherichia coli K12"
 /strain="K12"
 /sub_strain="MG1655"

```

repeat_region
<1..130
/note="REP (repetitive extragenic palindromic) element;
contains 4 REP sequences"
gene
/note="ybhc"
/note="b0772"
complement(147..1430)
CDS
/note="ybhc"
complement(147..1430)
/function="orf: Not classified"
/note="f427; 98 pct identical to fragment YBHC_ECOLI
SW:P46130 (300 aa) but contains 127 additional C-terminal
residues"
/codon_start=1
/transl_table=11
/product="putative pectinesterase"
/protein_id="AAC73859.1"
/db_xref="GI:1786989"
/translation="MNTFSVRLALALAFGVLTITACSTPPDQRPSPDQAPGTSRP1
LSAEQNPDAOHYFASLTGGAAMNPSPITLPADPVPVGPAGTQVHTTIOAAD
AALIKRTNKROYIAMPGEYQGVYVPAADGILYGTGEPIDVKIGLSLDGMSPA
DMRDVNPGRGMYGKPAWYMGSCOSKRSDSIGLCSAVFMSQNNGLQNLNTT
LGSVDAGNHPAVALRTDGDVOYIANNVILRGNTFEVNSGVQNLTEPQRTIVT
NSYTEGVIVSGGAVPNDTEPRVNSRTOOEAYVEAPATISNIYGLAANSRN
AFDGVAVQLGRSLVDVANTNGQVYTRDSAINESPTAKPWADAVISNRPAGTGSVD
DNDETQRNLMDTNNRMWENNRGVSRYAAEAKK"
promoter
complement(1507..1535)
/note="factor sigma70; predicted +1 start at 806574"
gene
complement(1582..2058)
/note="ybhb"
/note="b0773"
complement(1582..2058)
/function="orf: Unknown"
/note="f158; 99 pct identical to YBHB_ECOLI SW: P12994"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAC73860.1"
/db_xref="GI:1786990"
/translation="MKLISNDRDGDKLPHRHVNGNGYDGDNI SPHLANDVPAQTK
SFVVCYDPAPGSGMMHWVYVNLPADEFVLQGGSGGLVAMPDGLQTRIDPGTG
YDGAAPPKGETHRIYFVHALDIERIDVDGASGAMVGFVHPSLASASITAMFS"
protein_bind
/note="central position to predicted promoter:85.5"
/bound_moiety="Rhas predicted site"
complement(2117..3406)
gene
/note="b10A"
complement(2117..3406)
/gene="b10A"
/EC_number="2.6.1.62"
/function="enzyme: Biosynthesis of cofactors, carriers:
Biotin"
/note="f429; 100 pct identical to B10A_ECOLI SW: P12995"
/codon_start=1
/transl_table=11
/product="7,8-diaminopelargonic acid synthetase"
/db_xref="GI:1786991"
/translation="MTTDDIAFDQRIHMPYTSMTSELPYVYVSAEGCELLISDGR
LVQMSWMAIHGYNHQPLNAAMKSOIDMSVHVGITHAATIELCRKLVAMTPQ
LECVFLADSGVSEVAMKALQYQWAKGARORFLFRNGYDGFEGAMVCDPNS
MHSIMKGYLPENLFAPOSMDGEMDERMVFARIMAAHREIAVITEIPVQAG
GMRYVHEMLKRIKICDREGILIDAEITGCRGKLEACAEHAIAPDITICGAL
TGGMTLSATLTREVAETISNCEAGCGMPEMKPLACAAANSIATIESGDMCO
OVADIEVQLAEQIAPADAEVADYRVLGAYGVETTHPPVMAALQKFEVEGCVWIRP
FGKLTLPPTIILPQOLRQLTAAVNAVADDEFPCO"
promoter
complement(2193..2221)
/gene="b10A"
/note="factor sigma70; predicted +1 start at 807260"
3411..3450
/note="central position to b10b promoter: -20"

protein_bind
/bound_moiety="B10b documented site"
3411..3450
/note="central position to predicted promoter: -20; genetic
evidence for the site"
/bound_moiety="B10b predicted site"
complement(3411..3450)
/note="central position to predicted promoter:50"
/bound_moiety="B10b predicted site"
complement(3411..3450)
/note="central position to b10a promoter:50"
/bound_moiety="B10b documented site"
3413..3441
/note="factor sigma70; promoter b10b; documented +1 at
808525"
complement(3447..3473)
/note="factor sigma70; promoter b10a; documented +1 at
808515"
gene
3493..4533
/note="b10B"
/note="b0775"
3493..4533
/gene="b10B"
/EC_number="2.8.1.-"
/function="enzyme: Biosynthesis of cofactors, carriers:
Biotin"
/note="g346; 99 pct identical to B10B_ECOLI SW: P12996"
/codon_start=1
/transl_table=11
/product="biotin synthesis, sulfur insertion?"
/protein_id="AAC73862.1"
/db_xref="GI:1786992"
/translation="MAHRRWTLISQVTELEFKPLDLLFPAQYVHRQHFDPQOVST
LSTIKTGACPEDEKCYCPQSSRYKTGLEERLMEYQVESARKKAAGSTFCGAAM
KNHBERDMEYLEOMVOGYKAMGLEACMTLGLISEQALANAGIDYVNNMLDTPES
YONIITRTYQERLDTLEKVRDAGIKVSGIGVIGETVKRAGILLQALNLPPEP
VPIMLVKKVGPPLADNDVDADEFRTIARIMPTSVYVLSAGREONREORQAMC
FMAGANSYFGCKLITTPNPEDDKQLQFLRKLGNLPQOTAVLADNEDQQLBEALMT
PTDEYVNAAL"
4530..3684
/gene="b10F"
/note="b0776"
4530..3684
/gene="b10F"
/EC_number="2.3.1.47"
/function="enzyme: Biosynthesis of cofactors, carriers:
Biotin"
/note="g384; 100 pct identical to B10F_ECOLI SW: P12998"
/codon_start=1
/transl_table=11
/product="8-amino-7-oxononanoate synthase"
/protein_id="AAC73863.1"
/db_xref="GI:1786993"
/translation="MSWQEKINMALDARADALRRRYPAVQAGRGRLVADROYLNF
SSNDYLGSHHPQIIRAWOQGAEOFGTSSGSHVSGISVYHQALESLEALWLSRA
ILFISGFANQAVIAAMAKEDRIAADRLSASLSLEASLSPSQLRFAHNDVYHLR
ILASPQGOQAVTEGVFSSMDGSAPIAEIQOQVQOHNGMLVADHGTGYTGQGRG
SCWLQKVPBELLVTFYFGKFGVSGAVALCSSTVADYLLQFARHLITYSTMPAQAOL"

Query Match 100.0%; Score 25; DB 1; Length 11022;
Best local Similarity 100.0%; Pred. No. 0.0002;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcatgcacgattcgcctcgt 25
|||||
Db 3582 GCATGCCAGCATTCGATCCTCGT 3606

RESULT 13
AE005258 13501 bp DNA BCT 21-MAR-2001
LOCUS Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 82
DEFINITION of 155
ACCESSION AE005258 AE005174

```



```

VERSION      AE005258.1  GI:12513751
KEYWORDS
SOURCE
ORGANISM      Escherichia coli O157:H7 EDL933.
               Escherichia coli O157:H7 EDL933
               Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
               Escherichia.
REFERENCE
AUTHORS      1 (bases 1 to 13501)
               Perna, N.T., Plunkett, G., III, Burland, V., Mau, B., Glasner, J.D.,
               Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J., Kirkpatrick, H.A.,
               Postel, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L.,
               Grothbeck, E.J., Davis, N.W., Lim, A., Dimalanta, E., Potamousis, K.,
               Apodaca, J., Anantharaman, T.S., Lin, J., Yen, G., Schwartz, D.C.,
               Welch, R.A. and Blattner, F.R.
TITLE      Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
JOURNAL      Nature 409 (6819), 529-533 (2001)
MEDLINE      21074935
PUBMED      11206551
REFERENCE
AUTHORS      2 (bases 1 to 13501)
               Perna, N.T., Plunkett, G., III, Burland, V., Mau, B., Glasner, J.D.,
               Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J., Kirkpatrick, H.A.,
               Postel, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L.,
               Grothbeck, E.J., Davis, N.W., Lim, A., Dimalanta, E., Potamousis, K.,
               Apodaca, J., Anantharaman, T.S., Lin, J., Yen, G., Schwartz, D.C.,
               Welch, R.A. and Blattner, F.R.
TITLE      Direct Submission
JOURNAL      Submitted (22-OCT-2000) Laboratory of Genetics, University of
               Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
FEATURES
SOURCE
               1. 13501
               /organism="Escherichia coli O157:H7 EDL933"
               /strain="EDL933"
               /serotype="O157:H7"
               /db_xref="taxon:155864"
               /note="enterohemorrhagic"
               <1..7576
               /note="O-island #36; Region of the EDL933 chromosome not
               homologous to E. coli K-12 MG1655; CPElic prophage
               CP-933K; includes one copy of the 13 bp direct repeat that
               flanks the prophage"
               66..665
               /gene="lomk"
               /note="20981"
               66..665
               /gene="lomk"
               /function="putative membrane; Other or unknown (Phage or
               prophage Related)"
               /note="Residues 1 to 199 of 199 are 72.86 pct identical to
               residues 1 to 199 of 199 from Genpept 118 :
               gi175327891gb1AA63231.1|AF151091_2 (AF151091) lom
               [prophage P-B1bA]"
               /codon_start=1
               /transl_table=11
               /product="putative outer membrane protein of prophage
               CP-933K"
               /protein_id="AAG5137.1"
               /db_xref="GI:12513752"
               /translation="MRKVCALILSAICLANSGVPANASEHOSLTSLACYLHASTDAG
               SDDLNGIWKRYEFTDTLGLITSFSTANADECKTHSDTRMHEIDYVRNWSVMAG
               PSVRNENFSAVMAVAVSRVSTFGDYPVTNKRKTHDVLTHGSDDARYSNISLW
               GAGVOFNFTESAVADVAVEXSGSDMRTDGFIVGVYKF"
               719..2041
               /gene="20982"
               719..2041
               /gene="20982"
               /function="putative structure; Structural component (Phage
               or prophage Related)"
               /note="Residues 164 to 440 of 440 are 68.79 pct identical
               to residues 381 to 645 of 645 from Genpept 118 :
               gi145854361gb1AAD25464.1|AF125520.59 (AF125520) putative
               tail fiber protein [Bacteriophage 933W]"
               /codon_start=1
               /transl_table=11
               /product="putative tail component of prophage CP-933K"

gene
CDS
               /protein_id="AAG5138.1"
               /db_xref="GI:12513753"
               /translation="MGMAAYOISGVLRKDGAKGPIONCTIOLKARNSTXYVNTVASE
               NPDEGRISMDEYEQVSYTLVLVEGFPSPSHAGTISYDSDQPTLINFILGAMPEDDAR
               PEALRFPOWEEARHAEAKRNGEETSRNRNGISASAKBSANADTSADDAE
               SAROAEASAASAKKSEASSSASAEAAQASIESIADALISKTKTESAAGNARDA
               TSTTEKARESAQSAQSOIRIAEDAVNRIPVWPGPPGEGPAGPOGPKDGK
               RBDTPAGATGEGRGPGDTPAGPGPGAGDREGETLTGNAGPOGPGKGAAGA
               PGPGKPGKGAAGPYGATGPGPGKDDPETOIRFELGSMRIETNSYGMFPQTGALI
               TELTLDRDRAIQVQGMQHLQYVRGDPQMDVKDLDEGSDTGTGTE"
               2004..2312
               /gene="20984"
               2004..2312
               /gene="20984"
               /function="orf; Other or unknown (Phage or Prophage
               Related)"
               /note="Residues 1 to 102 of 102 are 98.03 pct identical to
               residues 28 to 129 of 129 from Genpept 118 :
               gi145854371gb1AAD25465.1|AF125520.60 (AF125520)
               hypothetical protein [Bacteriophage 933W]"
               /codon_start=1
               /transl_table=11
               /product="unknown protein encoded by prophage CP-933K"
               /protein_id="AAG5139.1"
               /db_xref="GI:12513754"
               /translation="MRKAVIOAELENDMNLRLKMQSLGCGSKHDCENGSLTQALR
               LCPADLLESDENGITPEQDRVITVOVILLADKKQIQCVVRPLQILRADGTWENIGMK"
               2489..3469
               /gene="20985"
               2489..3469
               /gene="20985"
               /function="orf; Other or unknown (Phage or Prophage
               Related)"
               /note="Residues 13 to 326 of 326 are 55.73 pct identical
               to residues 18 to 331 of 336 from Genpept 118 :
               gi169603671gb1AA63527.1 (AF170176) hypothetical protein
               predicted by Glimmer (Salmonella typhimurium LT2)"
               /codon_start=1
               /transl_table=11
               /product="unknown protein encoded by prophage CP-933K"
               /protein_id="AAG5140.1"
               /db_xref="GI:12513755"
               /translation="MLSPIRTFPHNSVINIYSSPCQVSPAGKEKELAVIDEKPILF
               OMFEPPPERYKKDEPVIYNTKQPIYDNTNARISSDMIGIPIVDGSPVQKTAFS
               KLERDENWVITIKREDVDPSMVKRKTSDIYHDICRQRLRTEDKRDREYLLNLEELR
               EISKADDSLSMTAKRKNAMDFEKNALILKAGEIFRCTYNTKNGISFPGGCIYD
               MDMLIGLGTITVAPDGISMDHVRNDSVNIETSAITVNSNPALBGLSPHMSKVD
               AHYPYGLGKGVKKYFNFTPLHNYNHCDFIEFNHPIIMNTSQTCSW"
               3530..4522
               /gene="20986"
               3530..4522
               /gene="20986"
               /function="orf; Other or unknown (Phage or Prophage
               Related)"
               /note="No significant matches"
               /codon_start=1
               /transl_table=11
               /product="unknown protein encoded by prophage CP-933K"
               /protein_id="AAG5141.1"
               /db_xref="GI:12513756"
               /translation="MKIPISLQSNFNSAPAGYSAPITAPNRANAVADYLDIGKRIPL
               SAADLSNVESEVIRAVHDSRSLIDQHTVIMIGNTVLDAISRSQTFRDVAVGIRINKE
               VHTGCIKYRNEYELNESSYKIDIDIOSLTJCNLEYDVGEPIFPLCEAGENDEEPY
               VRSVADPRTSPSYEMQBLIHHTVGHVSGDSDGSIDIEIGPELILARVAOELG
               WSPDFKGVAPEREALRLRNALRQAMRNHEENRAFEERLITISRYEASPPFT
               EYSAVSNIGYGFIOQHDPEGLAINDNLDANQIOLYHGAPYITTFQDVYKHNQO"
               5350..6231
               /gene="20989"
               5350..6231
               /gene="20989"
               /function="orf; Other or unknown (Phage or Prophage
               Related)"

```

```

/note="No significant matches"
/codon_start=1
/translation=1
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AG55142.1"
/db_xref="GI:12513757"
/translation="MSPSYVNLGCSMNSLTRITSPDNRLSSVRDAVHSDNGAV
KVGNNTRYVATDNKFCVTRSHSGCTNLHLRLGPKGEISRKIEVNLNAPSVAM
ERGIIVSNRDPDVPDVAPELPSVDNRSLSPGNVIGKGNNAVIEDADAKVLKM
FTTSQSEVTESEVRCFNOYAGSAAKIVGNNIDIGIMDKINGESLINSIPAO
AEHAIYDMFDEQKGLIPVDTETNVLVDRAKENFPIDSSINSDRSWSQSIMQ
SHGGKODLSIVLSVKI"
6462..7160
/gene="20990"
6462..7160
/gene="20990"
/function="orf; Other or unknown (Phage or Prophage
Related)"
/note="No significant matches"
/codon_start=1
/translation=1
/product="unknown protein encoded by prophage CP-933K"
/protein_id="AG55143.1"
/db_xref="GI:12513758"
/translation="MRPTSLNIVLHOSRSSMSDPIESLVASSYQWTKNPPQRF
OGDHNITQQIIEALDKTGTETGRVLLNATISIRKSETVITHNSSLGVMARD
IDAEHNGTGSDFHCNLAAYEPGEGISVVDHAIIVFHEHLHVFHNLNGERLAKES
SRASOKYSPLLLEARTVGLGAFSEVLSSENKHEIGMPRTSYPSXDSALIHDTN
VSLGFOVRLHPL"
complement(7668..8144)
/gene="ybhb"
/note="20992"
complement(7668..8144)
/gene="ybhb"
/function="orf; Unknown function"
/note="Residues 1 to 158 of 158 are 99.36 pct identical to
residues 1 to 158 of 158 from Escherichia coli K-12 Strain
MG1655: 80773"
/codon_start=1
/translation=1
/product="orf, hypothetical protein"
/protein_id="AG55144.1"
/db_xref="GI:12513759"
/translation="MKLISNDLRDGLKPLRHRYHNGMGYDGNISPHLAMDVPAGTK
SEVYCYDPAPTGSGMMHVVYVNLPAIDRVLPDGGSLVAMPQGVLOTRRDEKTC
YDCAAPKGETHRYITVHALDVERIDVDEGASGAVNGVNHFHSLASSTPAMS"
complement(8203..9492)
/gene="B10A"
/note="20993"
complement(8203..9492)
CDS
Query Match 100.0%; Score 25; DB 1; Length 13501;
Best local Similarity 100.0%; Pred. No. 0.00019;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gcatgcacagcattcgatcgcgt 25
|||||
Db 9668 GCATCGCGACGATTCGATCTCGT 9692
|||||
RESULT 14
AP002553 297816 bp DNA BCT 07-MAR-2001
LOCUS Escherichia coli O157:H7 DNA, complete genome, section 4/20.
DEFINITION AP002553 BA000007
ACCESSION AP002553.1 GI:13360211
KEYWORDS
SOURCE Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RIMD 0509952)
DNA.
ORGANISM Escherichia coli O157:H7
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1. (sites)

```

```

AUTHORS
Makino, K., Yokoyama, K., Kubota, Y., Yutsudo, C. H., Kimura, S.,
Kurokawa, K., Ishii, K., Hattori, M., Tatsuno, T., Abe, H., Iida, T.,
Yamamoto, K., Ohnishi, M., Hayashi, T., Yasunaga, T., Honda, T.,
Sasakawa, C. and Shinagawa, H.
JOURNAL
MEDLINE
REFERENCE
2 (sites)
Ohnishi, M., Murata, T., Nakayama, K., Kuhara, S., Hattori, M.,
Kurokawa, K., Yasunaga, T., Yokoyama, K., Makino, K., Shinagawa, H. and
Hayashi, T.
Comparative analysis of the whole set of rRNA operons between an
enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
Escherichia coli K-12 strain MG1655
Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
20557356
3 (sites)
Yokoyama, K., Makino, K., Kubota, Y., Watanabe, M., Kimura, S.,
Yutsudo, C. H., Kurokawa, K., Ishii, K., Hattori, M., Abe, H., Iida, T.,
Yamamoto, K., Hayashi, T., Yasunaga, T., Honda, T., Sasakawa, C. and
Shinagawa, H.
Complete nucleotide sequence of the prophage VT1-Sakai carrying the
Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
O157:H7 strain derived from the Sakai outbreak
Gene 258 (1-2), 127-139 (2000)
20564182
4 (sites)
Hayashi, T., Makino, K., Ohnishi, M., Kurokawa, K., Ishii, K.,
Yokoyama, K., Han, C.-G., Ohtsuno, E., Nakayama, K., Murata, T.,
Tanaka, M., Toke, T., Iida, T., Takami, H., Honda, T., Sasakawa, C.,
Ogasawara, N., Yasunaga, T., Kuhara, S., Shiba, T., Hattori, M. and
Shinagawa, H.
Complete genome sequence of enterohemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12
DNA Res. 8 (1), 11-22 (2001)
21156231
5 (bases 1 to 297816)
Ohnishi, M., Kurokawa, K., Makino, K., Yasunaga, T., Shinagawa, H. and
Hayashi, T.
Direct Submission
Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
Information Research Center, 3-1, Yamadaoka, Suita, Osaka 565-0871,
Japan (E-mail: ken@gen-info.osaka-u.ac.jp,
URL: http://www.gen-info.osaka-u.ac.jp/, Tel: 81-6-6879-8365,
Fax: 81-6-6879-2047)
genome project.
COMMENT
FEATURES
Location/Qualifiers
1..297816
/organism="Escherichia coli O157:H7"
/strain="O157:H7"
/sub_strain="RIMD 0509952"
/db_xref="taxon:83334"
79..1245
/gene="ECs0753"
79..1245
/gene="ECs0753"
/note="similar to SUCC_ECOLI g11786948 percent identity
100 in 388 aa (Conserved in E.coli K-12)"
/codon_start=1
/translation=1
/evidence="not experimental"
/product="succinyl-CoA synthetase beta subunit"
/protein_id="BAB34176.1"
/db_xref="GI:13360212"
/translation="MNLHEYOAKOLFAFGVLPAPGVYACTTPREAEASKIGAGPV
VKCOVHAGGGRKAGGVVYVNSKEDIPAFAPNMLGKRLVYQTPANQPNQIIVKAT
DIKELTIGAVYDRSSRRVYFMASTGGVEIEVVAETPHLHKVALDPLTGPYOG
RELAFLKGLGKLVQDFTKTEGLATFLERDLALIEINPLVTKGDLICIDGKIGA
DGNALFRQDPLREMRDQSDPREAQAQOMELVALDINIGCMVAGACILANGTMDIV
KLHGEPANFLDVGCGATKERTVFAFKIILSDKRVKAVLVNIFGGVIGVRCDLADGIG
AAVEGVNVPVVRLEGNNALGAKKLADSGNIIIAKGLTJAAQGVVAAYVEGK"

```


PUBMED 11133432
 2 (bases 1 to 5526)
 Entcheva, P., Liehl, W. and Streil, W. R.
 AUTHORS
 TITLE Direct Submission
 JOURNAL Submitted (31-MAR-2000) Mikrobiologie und Genetik, Universitaet
 Goettingen, Griesbachstr. 8, Goettingen 37077, Germany
 FEATURES
 source
 1..5526
 /organism="uncultured bacterium pCosHE2"
 /db_xref="taxon:143797"
 /clone="pCosHE2"
 /note="unknown organism, cosmid clone derived from
 environmental consortium"
 complement(52..528)
 /note="ORF1"
 /codon_start=1
 /transl_table=11
 /product="hypothetical 17.1 kDa protein in modC-bioA
 intergenic region"
 /protein_id="AAG60577.1"
 /db_xref="GI:12620125"
 /translation="MKLSNDLRDGRKLRHRYFNGMGVGDGNISPHLAMDVPAGTK
 SEVYTCYDPAEPTGSGMHWVNLPAIDRVLPQGGSLVAMPDGLQTRDPFGAG
 YDGAAPKGETHYIFTVALDIERIDVDEGASGAWGFNVHSHLASITAMES"
 complement(587..1876)
 /gene="bioA"
 complement(587..1876)
 /gene="bioA"
 /note="7.8-diaminopelargonic acid
 synthetase-aminotransferase"
 /codon_start=1
 /transl_table=11
 /product="DAPA-aminotransferase BioA"
 /protein_id="AAG60578.1"
 /db_xref="GI:12620126"
 /translation="MTTDLAFDQRIHHPYVSMTPSPVPSAEGCELLSDGR
 LVDSMSMAALHGVNHPQLNAKMSQIDAMSHVMGGITTHAPATELCRLVAMPQP
 LKCVLADSGSAVEAYAKMALQYQWAKGEARQPLTFENGHGFPGMSVCDSDNS
 MSLKMGYSPENLFAPOSRMDGEMDEMGFARLMAHHEHIAAVITEPIYGAG
 GMRMYHEMLKRIKIDREGILLTADETATGEGRGKLFACENHFIADILCIKAL
 TGGIMTLSATLTTRVEAETISNGEAGCFMGGTPGPNPLVACAAANSLATIESGMWO
 OVADIEVORLQRLARDAEMADVRLCATGVETTHFVNMAALQKFEVDEGVWIRP
 EKLITLMPYITLILQQLRTAAVNAVAVDETFFCQ"
 1963..3003
 /gene="bioB"
 1963..3003
 /gene="bioB"
 /codon_start=1
 /transl_table=11
 /product="biotin synthase BioB"
 /protein_id="AAG60579.1"
 /db_xref="GI:12620127"
 /translation="MAHRPMTLSQVTELEFEKPLDLLFEAQQVHRHOFDDROYVST
 LLSITGACPEPCQKPCOSSRYKIGLEAEPLMEVQYLSARAKAASSTFRCGAM
 KPNERNMPELEQYQWQKDLGELAEMLTGLTISBSQORLANAGLIDYINHLIDTSPER
 VFINLVKVGKPLADNDVDAFDFIRTLAARLIMPSTSVYLSNGRQNMNEQOAMC
 FMAGANSLEFGCKLLTTPNPEBKDLQLEFRKLGNPQQTAVLAGDNEQOQRLQALMT
 PPTDEYVNAAL"
 3000..4154
 /gene="bioF"
 3000..4154
 /gene="bioF"
 /note="8-amino-7-oxononanoate synthase"
 /codon_start=1
 /transl_table=11
 /product="KAPA synthetase BioF"
 /protein_id="AAG60580.1"
 /db_xref="GI:12620128"
 /translation="MSWOKTINALDAARAADALRRRYPAVAGARMLVADDDRYLNF
 SSNDYLGSHHPQITIRAWQGAQEQFGTSGSGSHVQSVVHQALFEELAEMLGYSRA
 LFTISGFAANQAVIAAMAKEDRIADRLSHASILEASLSPSOLRRFANHDVYHLAR
 LLASPCGOQLVTEGVFSMDGSAPLAEIQVITQGNHGLMVDANGTCVIGSGGNC

SCWLQKVPKELLVTFKGFVSGAVALCSSTVADYLLQFARHLIYSTMPPAQOAL
 RASLAVIRDEGDARREKLVSLIAHFRAGVODLPPTLADSCSAIOPLIVGNSRALQL
 AKLRQOCGWTAIRPPTVPAGTARLRLTLAAHMDIDILLETLEHONG"
 4141..4896
 /gene="bioc"
 4141..4896
 /gene="bioc"
 /note="reaction step prior to synthesis of pimeloyl-CoA"
 /codon_start=1
 /transl_table=11
 /product="biotin biosynthesis protein Bioc"
 /protein_id="AAG60581.1"
 /db_xref="GI:12620129"
 /translation="WATVNAKQIAAFAFGRAALYEOHADLQROSADYLLAMPKRYT
 HYLDAAGCGPMTTMRHREHQAUTALDISPMLVOAROKDAADHYLAGDIESPLATA
 TFDLMSNLAVQWGNLSTALRELYRVVRSQVAFETLVQSGSLPELHOAMQAVDERP
 HANRFLPDEIEOSLNGVYOHHTOPLITLWDDALSMRSKLGATGHIHGRDPRLL
 TRSQRLQRLAMPQOQGRPLTLHLFLGVARE"
 4883..5526
 /gene="biOD"
 4883..5526
 /gene="biOD"
 /note="PTB synthetase"
 /codon_start=1
 /transl_table=11
 /product="delthiobiotin synthetase BiOD"
 /protein_id="AAG60582.1"
 /db_xref="GI:12620130"
 /translation="MLVSKRYFVGTPTENGKTVASCALLQAKAAGARTAGKPYAS
 GSEKPEGLRNSDALQORNSLSLOLDATVAPYFAPRSPHITISNOEGPDISSVWS
 SGLRALROADMVIVBECAGCFPTLSOTFFPADVYTOQLPVLIVGVKLGICINHAWL
 TQALIQHAGLTLACVNAVADYTPPKRAEYITTLTMTAPALLEGITPMAENPE"
 BASE COUNT 1274 a 1507 c 1567 g 1178 t
 ORIGIN
 Query Match 96.0%; Score 24; DB 2; Length 5526;
 Best local similarity 100.0%; Pred. No. 0.00079;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 catgcacagcattcgatccctcgt 25
 Db 2053 CATGCCACGACATTGCGATCTCGT 2076
 RESULT 16
 ECOBIOB
 LOCUS 128 bp DNA BCT 03-JAN-1995
 DEFINITION Escherichia coli biotin (biob) gene, early terminator region.
 ACCESSION M27731
 VERSION M27731.1 GI:341755
 KEYWORDS biob gene; biotin.
 SOURCE Escherichia coli (strain K-12) DNA.
 ORGANISM Escherichia coli
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Escherichia.
 REFERENCE 1 (bases 1 to 128)
 AUTHORS Nath, S.K
 TITLE Attenuation of transcription of biotin genes in Escherichia coli
 JOURNAL Can. J. Microbiol. 34 (12), 1288-1296 (1988)
 MEDLINE 89167942
 FEATURES
 source
 1..128
 /organism="Escherichia coli"
 /strain="K-12"
 /db_xref="taxon:562"
 /gene="biob"
 4..128
 /gene="biob"
 4..9
 /gene="biob"
 20..128
 /gene="biob"
 103..121
 terminator
 -10_signal
 mRNA
 terminator

/gene="biog"
/note="early terminator"
BASE COUNT 31 a 28 c 34 g 35 t
ORIGIN

Query Match 72.0%; Score 18; DB 2; Length 128;
Best Local Similarity 100.0%; Pred. No. 3.2;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 8 cagattcgcattcgt 25
|||||
Db 57 CAGATTTCGATCTCCT 74

RESULT 17
G01446 479 bp DNA SFS 19-AUG-1999
LOCUS
DEFINITION Dm0466 Drosophila P1 library Drosophila melanogaster SFS genomic
clone DS07967 SP6, sequence tagged site.
ACCESSION G01446
VERSION G01446.1 GI:684849
KEYWORDS SFS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 479)
Rubin, G.
Berkeley Drosophila Genome Project
Unpublished (1994)
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Contact:
Berkeley Drosophila Genome Project
Primer A: CTCCTACCTGGGACATTCCTA
Primer B: CACCAAGGCGCATATCCGA
STS size: 192
PCR Profile:
Annealing: 58 degrees C PCR cycles: 30
Protocol:
Template: P1 Library Pools
Primer: 1 uM each
dNTPs: 250 uM each
Taq Poly: 0.05 units/uI
Total Vol: 15 uI
Buffer:
MgCl2: 1.5mM
KCl: 50 mM
Tris-HCl: 50 mM
pH: 8.3
Gelatin: .001 %

The P1 library has been distributed to 16 regional sites. A list
of these sites is available from FlyBase, via anonymous ftp to
ftp.bio.indiana.edu in the file
flybase/allied-data/genome-projects/1b1/LBL.doc.

FEATURES

source
1. 479
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="DS07967"
/note="Vector: Padiosacti; The P1 library was made by D.
Smoller in D. Hartl's lab (see Smoller et al., Chromosome
100: 487). Chromosomal position in the Hartl lab (see
Hartl et al., PNAS 91: 6824). SFS's were generated by
sequencing the ends of the Drosophila insert in these P1
clones, by the W. Kimmerly, C. Martin, and M. Palazzolo
lab at LBL."
1. .250

STS

primer_bind complement(233..250)
BASE COUNT 113 a 120 c 97 g 145 t 4 others
ORIGIN

Query Match 68.0%; Score 17; DB 54; Length 479;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gcatgccagcattcg 17
|||||
Db 396 GCATGCCAGCATTTGCG 412

RESULT 18
AC019940 53067 bp DNA HTG 03-JAN-2000
LOCUS
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS *** in ordered
pieces.
ACCESSION AC019940
VERSION AC019940.1 GI:6664957
KEYWORDS HTG; HTGS_PHASE1.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 53067)
Adams, M. and Venter, J.C.
Direct Submission
Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
This sequence was identified as CDM:10211463 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES

source
1. 53067
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"

BASE COUNT 14944 a 11879 c 11720 g 14524 t
ORIGIN

Query Match 68.0%; Score 17; DB 65; Length 53067;
Best Local Similarity 100.0%; Pred. No. 8.6;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcatgccagcattcg 17
|||||
Db 13629 GCATGCCAGCATTTGCG 13613

RESULT 19
AC013431 153870 bp DNA HTG 31-JAN-2000
LOCUS
DEFINITION Drosophila melanogaster chromosome X clone BACR22H11 (D1191)
RPCL-98 22.H.11 map 13E-13F strain y: cn bw sp, *** SEQUENCING IN
PROGRESS *** 73 unordered pieces.

ACCESSION AC013431
VERSION AC013431.8 GI:6838815
KEYWORDS HTG; HTGS_PHASE1.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 153870)
Celniker, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazek, R.G.,
Butenoff, C., Champagne, M., Chavez, C., Chew, M., Ciesiolka, L.,
Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,

TITLE
JOURNAL
AUTHORS
REFERENCE

Hinkle, A., Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K.,
Kearney, L., Lee, B., Lewis, S., Li, P., Ling, H., Moshireli, A.R.,
Moshireli, M., Nixon, K., Pacle, J.M., Park, S., Pfeiffer, B.,
Richards, S., Sethi, H., Svirska, R.R., Man, K.H., Webster, D.,
Woolley, P., Yang, S., Yee, M., Yu, C. and Rubin, G.M.
Sequencing of *Drosophila melanogaster*
Unpublished
2 (bases 1 to 153870)

TITLE
JOURNAL
COMMENT

Submitted (11-NOV-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Jan 31, 2000 this sequence version replaced q1:6532028.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (<http://www.fruitfly.org/sequence/>) or send email
to hdg@fruitfly.berkeley.edu. All contigs in this submission meet
the following cutoffs: length >= 200 bases.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 73 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 995: contig of 995 bp in length
996 1075: gap of unknown length
1076 1560: contig of 485 bp in length
1561 1640: gap of unknown length
1641 2230: contig of 590 bp in length
2231 2310: gap of unknown length
2311 2956: contig of 646 bp in length
2957 3036: gap of unknown length
3037 3621: contig of 585 bp in length
3622 3701: gap of unknown length
3702 4535: contig of 834 bp in length
4536 4615: gap of unknown length
4616 5559: contig of 944 bp in length
5560 5639: gap of unknown length
5640 6535: gap of 896 bp in length
6536 6615: gap of unknown length
6616 7660: contig of 1045 bp in length
7661 7740: gap of unknown length
7741 8730: contig of 990 bp in length
8731 8810: gap of unknown length
8811 9404: contig of 594 bp in length
9405 9484: gap of unknown length
9485 10725: contig of 1241 bp in length
10726 10805: gap of unknown length
10806 11578: contig of 773 bp in length
11579 11658: gap of unknown length
11659 12834: contig of 1176 bp in length
12835 12914: gap of unknown length
12915 13501: contig of 587 bp in length
13502 13581: gap of unknown length
13582 14462: contig of 881 bp in length
14463 14542: gap of unknown length
14543 15566: contig of 1024 bp in length
15567 15646: gap of unknown length
15647 16366: contig of 720 bp in length
16367 16446: gap of unknown length
16447 17301: contig of 855 bp in length
17302 17381: gap of unknown length
17382 18390: contig of 1009 bp in length
18391 18470: gap of unknown length

18471 19409: contig of 939 bp in length
19410 19489: gap of unknown length
19490 20726: contig of 1237 bp in length
20727 20806: gap of unknown length
20807 22052: contig of 1246 bp in length
22053 22132: gap of unknown length
22133 23015: contig of 883 bp in length
23016 23095: gap of unknown length
23096 24523: contig of 1527 bp in length
24523 24702: gap of unknown length
24703 26689: contig of 1987 bp in length
26690 26769: gap of unknown length
26770 27825: contig of 1056 bp in length
27826 27905: gap of unknown length
27906 29904: contig of 1999 bp in length
29905 29984: gap of unknown length
29985 31588: contig of 1604 bp in length
31589 31668: gap of unknown length
31669 33112: contig of 1444 bp in length
33113 33192: gap of unknown length
33193 35374: contig of 2182 bp in length
35375 35454: gap of unknown length
35455 37223: contig of 1769 bp in length
37224 37303: gap of unknown length
37304 38978: contig of 1675 bp in length
38979 39058: gap of unknown length
39059 41337: contig of 2279 bp in length
41338 41417: gap of unknown length
41418 42461: contig of 1044 bp in length
42462 42561: gap of unknown length
42562 45823: contig of 3282 bp in length
45824 45903: gap of unknown length
45904 47247: contig of 1344 bp in length
47248 47327: gap of unknown length
47328 50153: contig of 2826 bp in length
50154 50233: gap of unknown length
50234 52182: contig of 1949 bp in length
52183 52262: gap of unknown length
52263 56756: contig of 4494 bp in length
56757 56836: gap of unknown length
56837 63525: contig of 6689 bp in length
63526 63605: gap of unknown length
63606 71374: contig of 7669 bp in length
71375 71354: gap of unknown length
71355 78141: contig of 6787 bp in length
78142 78221: gap of unknown length
78222 86570: contig of 8349 bp in length
86571 86650: gap of unknown length
86651 96227: contig of 9577 bp in length
96228 96307: gap of unknown length
96308 107431: contig of 11124 bp in length
107432 107511: gap of unknown length
107512 121618: contig of 14107 bp in length
121619 121698: gap of unknown length
121699 139797: contig of 18099 bp in length
139798 139877: gap of unknown length
139878 140388: contig of 511 bp in length
140389 140468: gap of unknown length
140469 141011: contig of 543 bp in length
141012 141091: gap of unknown length
141092 141513: contig of 422 bp in length
141514 141593: gap of unknown length
141594 142057: contig of 464 bp in length
142058 142137: gap of unknown length
142138 142690: contig of 553 bp in length
142691 142770: gap of unknown length
142771 143344: contig of 574 bp in length
143345 143424: gap of unknown length
143425 143857: contig of 433 bp in length
143858 143937: gap of unknown length
143938 144603: contig of 666 bp in length
144604 144683: gap of unknown length
144684 145116: contig of 433 bp in length

```

* 145117 145196: gap of unknown length
* 145197 145734: contig of 538 bp in length
* 145735 145814: gap of unknown length
* 145815 146224: contig of 410 bp in length
* 146225 146304: gap of unknown length
* 146305 146859: contig of 555 bp in length
* 146860 146939: gap of unknown length
* 146940 147585: contig of 646 bp in length
* 147586 147665: gap of unknown length
* 147666 148081: contig of 416 bp in length
* 148082 148161: gap of unknown length
* 148162 148539: contig of 378 bp in length
* 148540 148619: gap of unknown length
* 148620 148950: contig of 331 bp in length
* 148951 149030: gap of unknown length
* 149031 149375: contig of 345 bp in length
* 149376 149455: gap of unknown length
* 149456 149853: contig of 398 bp in length
* 149854 149933: gap of unknown length
* 150532 150532: contig of 599 bp in length
* 150533 150612: gap of unknown length
* 150613 151147: contig of 535 bp in length
* 151148 151227: gap of unknown length
* 151228 151577: contig of 350 bp in length
* 151578 151657: gap of unknown length
* 151658 151964: contig of 307 bp in length
* 151965 152044: gap of unknown length
* 152045 152629: contig of 585 bp in length
* 152630 152709: gap of unknown length
* 152710 153245: contig of 536 bp in length
* 153246 153325: gap of unknown length
* 153326 153870: contig of 545 bp in length.

```

```

FEATURES
  source
    1. 153870
      /organism="Drosophila melanogaster"
      /strain="Y: cn bw sp"
      /db_xref="taxon:7227"
      /chromosome="X"

```

```

Query Match      68.0%; Score 17; DB 63; Length 153870;
Best Local Similarity 100.0%; Pred. No. 8.1;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 gcatcgccagcattcg 17
|||||
Db 63673 GCATCGCCAGCATTCG 63689

```

```

RESULT 20
AC007809 183439 bp DNA INV 24-FEB-2001
LOCUS Drosophila melanogaster, chromosome 3R, region 88C-88C, BAC clone
DEFINITION BACR45M03, complete sequence.
ACCESSION AC007809
VERSION AC007809.8 GI:13122709
KEYWORDS HTG.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
          Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
          Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
          Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
          1 (bases 1 to 183439)
REFERENCE 1
  Celniker,S.E., Adams,M.D., Krommiller,B., Tyler,D., Wan,K.H.,
  Holt,R.A., Evans,C.A., Gocayne,J.D., Amanalides,P.G., Brandon,R.C.,
  Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
  Carlson,J.W., Center,A., Champs,M., Davenport,L.B., Dietz,S.M.,
  Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
  Ferreira,S., Frise,E., Galle,R.F., Gary,N.S., George,R.A.,
  Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howland,T.J.,
  Ilegwam,C., Jalili,M., Kruse,D., Li,P., Matelli,B., Moshrefi,A.,
  McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nunoo,J.,
  Paiele,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
  Phouanavong,S., Pittman,G.S., Puri,V., Richards,S., Scheeler,F.,

```

```

TITLE Stapleton,M., Strong,R., Svirkas,R., Tector,C., Williams,S.M.,
JOURNAL Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
REFERENCE Sequencing of Drosophila chromosome 3R, region 88C-88C
  Unpublished
  2 (bases 1 to 183439)
AUTHORS Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
  Butenhoff,C., Champs,M., Chavez,C., Chew,M., Ciesiolka,L.,
  Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
  Hoskins,R.A., Houston,K.A., Hummel,S.R., Karra,K., Kearney,L.,
  Kim,E., Lee,B., Lewis,S., Li,P., Lomtan,M.A., Mazda,P.,
  Moshrefi,A.R., Moshrefi,M., Nixon,K., Paiele,J.M., Park,S.,
  Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
  Svirkas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
  Rubin,G.M.

```

```

COMMENT Direct Submission
  Submitted (14-JUN-1999) Drosophila Genome Center, Lawrence Berkeley
  Laboratory, MS 64-121, Berkeley, CA 94720, USA
  On Feb 24, 2001 this sequence version replaced g1:6623906.
  Sequence submitted by:
  Berkeley Drosophila Genome Project
  Lawrence Berkeley National Laboratory, MS 64-121
  Berkeley, CA 94720

```

```

This sequence was assembled using end sequences from a whole genome
shotgun and from subclones of this BAC and its neighboring clones.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (http://www.fruitfly.org/sequence/) or send email
to bdgp@fruitfly.berkeley.edu.

```

```

FEATURES
  source
    1. 183439
      /organism="Drosophila melanogaster"
      /strain="Y: cn bw sp"
      /db_xref="taxon:7227"
      /chromosome="3R"
      /map="88C-88C"
      /clone="BACR45M03 (D718)"
      /clone_id="RPCT-98 (Roswell Park Cancer Institute
      Drosophila melanogaster BAC library, partial EcoRI in
      PBAC3.6)"

```

```

BASE COUNT 53300 a 38849 c 38854 g 52436 t
ORIGIN

```

```

Query Match      68.0%; Score 17; DB 4; Length 183439;
Best Local Similarity 100.0%; Pred. No. 8;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 4 tcgcccagcattcgatc 20
|||||
Db 66707 TCGCCAGCATTCGATC 66723

```

```

RESULT 21
AE003705 232744 bp DNA INV 05-OCT-2000
LOCUS Drosophila melanogaster genomic scaffold 142000013386035 section 30
DEFINITION of 105, complete sequence.
ACCESSION AE003705 AE002708
VERSION AE003705.1 GI:7299886
KEYWORDS HTG.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
          Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
          Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
          Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
          1 (bases 1 to 232744)
REFERENCE 1
  Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
  Amanalides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
  George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
  Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
  Britton,R.C., Rogers,J.H., Blazej,R.G., Champs,M., Pfeiffer,B.D.,
  Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gaber
  Miklos,G.L., Adril,J.F., Agbayani,A., An,H.J.,

```

Andrews-Pfannkuch,C., Baldwin,D., Ballew,R.M., Basu,A.,
 Baxendale,J., Bayraktaroglu,I., Beasley,E.M., Beeson,K.Y.,
 Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
 Botcharov,M.R., Bouck,J., Brokstein,P., Brotlier,P., Burtils,K.C.,
 Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
 Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
 Pablo,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
 Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
 Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferreira,S.,
 Fleischmann,W., Foster,C., Gabrielian,A.E., Garg,N.S.,
 Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
 Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
 Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
 Wei,M.H., Idegwam,C., Jalali,M., Kalush,F., Karpn,G.H., Ke,Z.,
 Krenitsen,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
 Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
 Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
 McLeod,M.P., McPherson,D., Mekulov,G., Mishina,N.V., Mobarry,C.,
 Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
 Murny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
 Nusskern,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
 Pollard,D., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
 Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
 Stapleton,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
 Stetler,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,
 Weinstein,G.M., Weissenbach,J., Williams,S.M., Woodage,T.,
 Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
 Zaveri,J.S., Zhang,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
 Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
 Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
 The genome sequence of *Drosophila melanogaster*
 Science 287 (5461), 2185-2195 (2000)

2 (bases 1 to 232744)
 Adams,M.D., Celisner,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
 Direct Submission
 Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
 Rockville, MD, USA

FEATURES

Source
 Location/Qualifiers
 1..232744
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /chromosome="3R"
 complement(join(29274..29439,29524..30038,30107..>30980))
 /gene="CG8464"
 /product="CT24781"
 /db_xref="FLYBASE:FBan0008464"
 /db_xref="FLYBASE:FBgn0038233"
 complement(<29274..>30980)
 /gene="CG8464"
 /map="88C4-88C6"
 /db_xref="FLYBASE:FBan0008464"
 /db_xref="FLYBASE:FBgn0038233"
 complement(join(29274..29439,29524..30038,30107..30694))
 /gene="CG8464"
 /note="CG8464 gene product"
 /codon_start=1
 /db_xref="FLYBASE:FBan0008464"
 /db_xref="FLYBASE:FBgn0038233"
 /protein_id="AAF5062.1"
 /db_xref="GI:7299887"
 /translation="MALRGSRLLEVIFKRCIASPVLSHAANRSSLAIKRGDPNSN
 GNSGYOQNGEOKERKMRVLFEPVSLGAVSAIIQREDDLPITIAASKWTCGRD
 FNEIADYVAGCADSVYIEIKDRHDFVSGPIITANSNGGITFONCLILNNAIVY
 NKPHTMVQRYLSDGRFPATIEDVQTSGLATLRQVNNLSMRIGKSTLRSGEYV
 AGSPILASNTYTAGVISTQASQELGRNNDIYDTDAITFGNSGGPLVNLDE
 AIGVSMKVTAGISFAIPIDYKVELERAEKRRKGSAYKGYPKRMYGIMTLITP
 DILFEIKSRQMPNSLTHGVLMKVIVGSPHNSGLOPCDVIHINKKEIKNSDVS
 DALANSKTLDIVILRGVOMHVTITPEDP"
 join(<31212..31346,31407..31691,31752..>31922)
 /gene="CG3351"
 /product="CT11255"

gene
 /db_xref="FLYBASE:FBan0003351"
 /db_xref="FLYBASE:FBgn0038234"
 <31212..>31922
 /gene="CG3351"
 /map="88C4-88C6"
 /db_xref="FLYBASE:FBan0003351"
 /db_xref="FLYBASE:FBgn0038234"
 join(31212..31346,31407..31691,31752..31922)
 /gene="CG3351"
 /note="CG3351 gene product"
 /codon_start=1
 /db_xref="FLYBASE:FBan0003351"
 /db_xref="FLYBASE:FBgn0038234"
 /protein_id="AAF5063.1"
 /db_xref="GI:7299888"
 /translation="MSKAAGKLSLKTVERVYHTSKLTNIPAGMAAGPLGPMIG
 ORAINIAFCQDNFAKTBEMKGVPLPQRISSNDRSYDLAIHPDPATFFLKAGIO
 RGTMTGKEVAGMITKLHLEYETAIKIDPPNVLLTQWOCMEMLISTARTGIVYRE
 IDPAVGEFLERKLIQOORREIQERKARMLRTG"
 complement(join(32136..32708,32770..>32889))
 /gene="CG8461"
 /product="CT24741"
 /db_xref="FLYBASE:FBan0008461"
 /db_xref="FLYBASE:FBgn0038235"
 complement(<32136..>32889)
 /gene="CG8461"
 /map="88C4-88C6"
 /db_xref="FLYBASE:FBan0008461"
 /db_xref="FLYBASE:FBgn0038235"
 complement(join(32136..32708,32770..32835))
 /gene="CG8461"
 /note="CG8461 gene product"
 /codon_start=1
 /db_xref="FLYBASE:FBan0008461"
 /db_xref="FLYBASE:FBgn0038235"
 /protein_id="AAF5064.1"
 /db_xref="GI:7299889"
 /translation="MSSPSSSSASSEHSDNDNETONAIREDLKGSFEEIMLKLEEL
 GAKYKEAVILGSKSSRPQPKAKTDLRLKLNRPREMTTRQVPEFLAEHRVSEKKVY
 ELRDPREKESGTYSAEETPKKQYQFVSRKEVGLKKLDEVEDAEKHHIKDTMIO
 RLKNNVEDKKWHKOKOLKREKRSALIKHDLGOOPHYLRKSPFLIDSEIKF"
 complement(join(<34919..35540,35598..35757,35894..36047,
 36334..36679,36886..37166,37233..37324,37381..37624,
 37694..37981,38058..38138,42442..42707,42765..43010,
 44497..44761,48465..>49118))
 /gene="CG8279"
 /product="CT24224"
 /db_xref="FLYBASE:FBan0008279"
 /db_xref="FLYBASE:FBgn0038237"
 complement(<34919..>49118)
 /gene="CG8279"
 /map="88C5-88C6"
 /db_xref="FLYBASE:FBan0008279"
 /db_xref="FLYBASE:FBgn0038237"
 complement(join(34919..35540,35598..35757,35894..36047,
 36334..36679,36886..37166,37233..37324,37381..37624,
 37694..37981,38058..38138,42442..42707,42765..43010,
 44497..44761,48465..49118))
 /gene="CG8279"
 /note="CG8279 gene product"
 /codon_start=1
 /db_xref="FLYBASE:FBan0008279"
 /db_xref="FLYBASE:FBgn0038237"
 /protein_id="AAF5066.2"
 /db_xref="GI:10726529"
 /translation="MHGPVSRSSSSNMVDVSSPAGGAAPVEMSTTSSSAAATTSAS
 SSKPLTNGAKMTAISTAGAGVTPGAVGPCSGAIPASSSSGNOVKLEHHRVSEKNNRP
 AVNRSETTILMTPTGSSSPQSPQSTQASIOIOTQSDRLKAKSTTASQODVDA
 RLPEERKEAPEKTLERAPPEALSRIQRIENKPKRRSVSDLEQQMAASPTVQO
 KSRFSLSNASSSLPECRHMDLDEGELEMLIRQVANELEDIVLCHILVAVGLT
 HADRSGLFLAKGPTNKKYIVAKLELDVQTKALADATYRASAEIILIPGIGIAGVAQO
 TKOMINIKAVKDAFENCETIDLTGTGTNAALICMPQNCNEDGIDIGVAQIINTNGME
 FDEHDEIEFRRLTFPGIGIQNQLFEMSVQEFYRRNQITLLNARSIFEDQNNDECLVT

KIMTARELLKCCRSVFLVDLCCASHLEKIEKPNOPATRAIKSADSEEEKMRN
 RETVLFEIGEGEQAANVRSPVSELSSSTLQIAOFVATGCTVACIVIEVRNHD
 IRAEDFIDSTOAILCIMPIMNAKQVIGVAGOLINKANGVPTDSDAIEFAPGGLG
 JINTOMYENACKLMAKOKVALECLSYHATASODPOTEKLTODVIAESNYSFPTD
 FELVDDDTCAVLRMRMOCNIVSOPFIPDYLCRVNLSVRKRYKPRHNRHNLNVA
 QTFMALKTKMKMERMTLEILGLVACICHDHGRGNMVEQITESPPLILYTTST
 MEHNDHFRILAFILRLAARALRLIARLIDVIELVNDPILFVIEFRLLR
 TLEFPELEDEDDVDVSVVSVLVLDVSVLILSALTWATYKTIERTKRT
 RDNFMDALSPEDYSKMTVESALISDLDLAEFKRNALFELVNGSEFDMGEEKRD
 LLCGMATACDVSAIAKPRVVOHAKVLADEFPPOGDEKLTQITOPVAMDRERKD
 ELPRKMGVFDIVICLPYLVLCDFPWTPIYEGTLVGNRMODLAEKVEKLTMDIDH
 DTIDRVEFEFFACADEKIDIEFTVTTLNCSQSHGSEDSITPHOSGSLSKMT
 GALGAVRSLKSLTYLNSMDGSKPKTSKLLESHVSEDMDKSPSPQPOASGSMGR
 MSASSTSSAGGQMDKSKRSKLCALL"
 join(40359, 40547, 40626, 40826, 40891, 41076, 41138, 41210,
 41284, 41410, 41461, 42034, 42102, 42262)
 /gene="Cyp313a1"
 /product="Cp313a1"
 /db_xref="FLYBASE:FBgn0003360"
 /db_xref="FLYBASE:FBgn0038236"
 <40359, >42262
 /gene="Cyp313a1"
 /note="CG3360"
 /map="88C5-88C6"
 /db_xref="FLYBASE:FBgn0003360"
 /db_xref="FLYBASE:FBgn0038236"
 join(40391, 40547, 40626, 40826, 40891, 41076, 41138, 41210,
 41284, 41410, 41461, 42034, 42102, 42262)
 /gene="Cyp313a1"
 /note="Cyp313a1 gene product"
 /codon_start=1
 /db_xref="FLYBASE:FBgn0003360"
 /db_xref="FLYBASE:FBgn0038236"
 /protein_id="AAF55065.2"
 /db_xref="GI:10726530"

Query Match 68.0%; Score 17; DB 5; Length 232744;
 Best Local Similarity 100.0%; Pctd. No. 7.9;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 tgcgcagcattcgatc 20
 |||||
 Db 106438 TCGCAGCATTTCGATC 106454

RESULT 22
 AC017740/c DNA HTG 10-DEC-1999
 LOCUS
 DEFINITION AC017740 237119 bp
 AC017740 Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***
 VERSION AC017740.1 GI:6554259
 KEYWORDS HTG; HTGS_PHASE2.
 SOURCE fruit fly.
 ORGANISM Drosophila melanogaster
 Drosophila: Arthropoda; Tracheata; Hexapoda; Insecta;
 Eukaryota; Metazoa; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 237119)
 Adams, M. and Venter, J.C.
 Direct Submission
 Submitted (10-DEC-1999) Celera Genomics, 45 West Gude Drive,
 Rockville, MD, USA
 COMMENT This sequence was identified as CDM:10211941 by the submitter.
 For more information on this record e-mail to fly@celera.com.
 * NOTE: This is a 'working draft' sequence.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 Location/Qualifiers
 1..237119
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"

BASE COUNT 67689 a 49909 c 50172 g 69349 t
 ORIGIN
 Query Match 68.0%; Score 17; DB 65; Length 237119;
 Best Local Similarity 100.0%; Pctd. No. 7.9;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 tgcgcagcattcgatc 20
 |||||
 Db 103681 TCGCAGCATTTCGATC 103665

RESULT 23
 AE003500 INV 04-OCN-2000
 LOCUS Drosophila melanogaster genomic scaffold 142000013386053 section 17
 DEFINITION of 30, complete sequence.
 ACCESSION AE003500 AE002593
 VERSION AE003500.2 GI:10728273
 KEYWORDS HTG.
 SOURCE fruit fly.
 ORGANISM Drosophila melanogaster
 Drosophila: Arthropoda; Tracheata; Hexapoda; Insecta;
 Eukaryota; Metazoa; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 327446)
 Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D.,
 Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Gallie, R.F.,
 George, R.A., Lewis, S.E., Richards, S., Ashburner, M., Henderson, S.N.,
 Sutton, G.G., Wortman, J.R., Yandell, M.D., Zhang, Q., Chen, L.X.,
 Burton, R.C., Rogers, Y.H., Blazer, R.G., Chapple, M., Pfeiffer, B.D.,
 Wan, K.H., Doyle, C., Baxter, E.G., Helt, G., Nelson, C.R., Gabor, M.,
 Miklos, G.L., Abril, J.F., Agbayani, A., An, H.J.,
 Andrews, P., Binkov, C., Baldwin, D., Balow, R.M., Basu, A.,
 Baxendale, J., Bayraktaroglu, L., Beasley, E.M., Beeson, K.Y.,
 Benos, P.V., Bernier, B.P., Bhandari, D., Bolshakov, S., Borokov, B.C.,
 Botchan, M.A., Bouck, J., Brokstein, P., Brotlier, P., Burris, K.C.,
 Busam, D.A., Butler, H., Cadieu, E., Center, A., Chandra, I.,
 Cherry, J.M., Cawley, S., Dahlke, C., Davenport, L.B., Davies, P., de
 Pablo, B., Delcher, A., Deng, Z., Mays, A.D., Dew, I., Dietz, S.M.,
 Dodson, K., Dou, L.E., Downes, M., Dugan-Rocha, S., Dunkov, B.C.,
 Dunn, P., Durbin, K.J., Evangelista, C.C., Ferraz, C., Ferrier, S.,
 Fleischmann, W., Foster, C., Gabrielian, A.E., Gary, N.S.,
 Gelbart, W.M., Glasser, K., Glodek, A., Gong, F., Gorelli, J.H., Gu, Z.,
 Guan, P., Harris, M., Harris, N.L., Harvey, D., Heiman, T.J.,
 Hernandez, J.R., Houck, J., Hostin, D., Houston, K.A., Howland, T.J.,
 Wel, M.H., Ibegwam, C., Jalali, M., Kalush, F., Karpen, G.H., Ke, Z.,
 Kennison, J.A., Ketchum, K.A., Kimmel, B.E., Kodira, C.D., Kraft, C.,
 Kravitz, S., Kulp, D., Lai, Z., Lasko, P., Lei, Y., Levitsky, A.A.,
 Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Mattei, B., McIntosh, T.C.,
 McLeod, M.P., McPherson, D., Merkulov, G., Mishina, N.V., Mobarry, C.,
 Morris, J., Moshrefi, A., Mount, S.M., Moy, M., Murphy, B., Murphy, L.,
 Muzny, D.M., Nelson, D.L., Nelson, D.R., Nelson, K.A., Nixon, K.,
 Nusskern, D.R., Pacle, J.M., Palazolo, M., Peltman, G.S., Pan, S.,
 Pollard, J.D., Puri, V., Reese, M.G., Reiter, K., Remington, K.,
 Saunders, R.D., Scheeler, F., Shen, H., Shue, B.C., Siden-Kiamos, I.,
 Simpson, M., Skupski, M.P., Smith, T., Spier, E., Spradling, A.C.,
 Stapleton, M., Strong, R., Sun, E., Svatek, R., Tector, C., Turner, R.,
 Venter, E., Wang, A.H., Wang, X., Wang, Z.Y., Wasserman, D.A.,
 Weinstein, G.M., Weissenbach, J., Williams, S.M., Woodage, T.,
 Worley, K.C., Wu, D., Yang, S., Yao, Q.A., Ye, J., Yeh, R.F.,
 Zaveri, J.S., Zhang, M., Zhang, S., Zhao, Q., Zheng, L., Zheng, X.H.,
 Zhong, F.N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H.O.,
 Gibbs, R.A., Myers, E.W., Rubin, G.M. and Venter, J.C.
 The genome sequence of Drosophila melanogaster
 Science 287 (5461), 2185-2195 (2000)
 20196006
 2 (bases 1 to 327446)
 Adams, M.D., Celniker, S.E., Gibbs, R.A., Rubin, G.M. and Venter, J.C.
 Direct Submission
 Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
 Rockville, MD, USA

OY 1 gcatgccagcatttcg 17
 Db 70808 GCATGCCAGCATTTCG 70824
 RESULT 24
 AF146614/c 1186 bp DNA BCT 09-AUG-1999
 LOCUS
 DEFINITION Erwinia carotovora subsp. carotovora xerD (xerD) gene, partial cds;
 DsbC precursor (dsbC) gene, complete cds; and RecJ (recJ) gene,
 partial cds.
 ACCESSION AF146614 GI:5712695
 VERSION
 KEYWORDS Pectobacterium carotovorum subsp. carotovorum,
 Pectobacterium carotovorum subsp. carotovorum
 ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 Pectobacterium
 REFERENCE 1 (bases 1 to 1186)
 AUTHORS Vincent-Sealy,L.V., Thomas,J.D., Commander,P. and Salmond,G.P.
 TITLE Erwinia carotovora Dsba mutants: evidence for a periplasmic-stress
 signal transduction system affecting transcription of genes
 encoding secreted proteins
 JOURNAL Microbiology 145 (Pt 8), 1945-1958 (1999)
 MEDLINE 99392457
 REFERENCES 2 (bases 1 to 1186)
 AUTHORS Thomas,J.D.
 TITLE Direct Submission
 JOURNAL Submitted (29-APR-1999) Biological Sciences, University of Warwick,
 Gibbet Hill Road, Coventry CV4 7AL, UK
 FEATURES
 Source
 1..1186
 /organism="Pectobacterium carotovorum subsp. carotovorum"
 /strain="SCRI193"
 /sub_species="carotovora"
 /db_xref="taxon:555"
 <1..189
 /gene="xerD"
 <1..189
 /gene="xerD"
 /note="similar to Escherichia coli xerD
 integrase/recombinase in SwissProt Accession Number
 P21891"
 /codon_start=1
 /transl_table=11
 /evidence=not_experimental
 /product="xerD"
 /protein_id="AAD47611.1"
 /db_xref="GI:5712697"
 /translation="KLSPHYLRHAFATHLLNHGAGFTGRDLAGQKSPFDADLHPCR
 DGAETRASAHRHRTT"
 222..1031
 /gene="dsbC"
 222..227
 /gene="dsbC"
 246..251
 /gene="dsbC"
 258..264
 /gene="dsbC"
 /note="putative"
 315..377
 /gene="dsbC"
 /note="putative sec-dependent export signal"
 315..1031
 /gene="dsbC"
 /function="periplasmic disulfide bond formation"
 /note="disulfide isomerase; similar to Erwinia
 chrysanthemi and Escherichia coli DsbC proteins in
 SwissProt Accession Numbers P39691 and P21892"
 /codon_start=1
 /transl_table=11
 /product="DsbC precursor"
 CDS
 sig_peptide

mat_peptide
 misc_feature
 gene
 CDS
 /protein_id="AAD47610.1"
 /db_xref="GI:5712696"
 /translation="MKKGLISLVATATNFHADAATIKOTLHAGVGARGGPAPA
 DGRKTVITDSDGLYISDGGAPARAPLRERTRPVNTNKKILGDLNLOEMIV
 KAAQERHVTFTSTIGYCHRLKEDMDYALGIVRYLAFPRQKSPRAKQMSI
 MCVADRNRAPDSAMKGAISAATCKTDIAAHYQLGIDGVOGTALVLDGLVPGIO
 GPKEMLAMIEAHKASKRTGG"
 378..1028
 /gene="dsbC"
 /product="DsbC"
 669..680
 /gene="dsbC"
 /note="Region: putative CPHC motif typical of the active
 sites of disulfide isomerases"
 1044..>1186
 /gene="recJ"
 1044..>1186
 /gene="recJ"
 /note="theoretical protein; similar to Erwinia
 chrysanthemi and Escherichia coli RecJ single-stranded
 DNA-specific exonucleases in SwissProt Accession Numbers
 P39693 and P21893"
 /codon_start=1
 /transl_table=11
 /evidence=not_experimental
 /product="RecJ"
 /protein_id="AAD47612.1"
 /db_xref="GI:5712698"
 /translation="MITQLRRRPLAEDIDLPVTVPPLRLRYAHGKVGKQAELESL
 GL"
 BASE COUNT 295 a 315 c 323 g 253 t
 ORIGIN
 Query Match 64.0%; Score 16; DB 2; Length 1186;
 Best Local Similarity 100.0%; Pred. No. 41;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 gcatgccagcatttc 16
 Db 993 GCATGCCAGCATTTC 978
 RESULT 25
 AF206717/c 11910 bp DNA SYN 12-SEP-2000
 LOCUS
 DEFINITION Shuttle vector p13 hypothetical proteins, Htrv (htrv), Ardu (ardu),
 Repu (repu), and Resu (resu) genes, complete cds.
 ACCESSION AF206717
 VERSION AF206717.1 GI:7262998
 KEYWORDS
 SOURCE Shuttle vector p13.
 ORGANISM Shuttle vector p13.
 REFERENCE 1 (bases 1 to 11910)
 AUTHORS Meima,R. and Lidstrom,M.E.
 TITLE Characterization of the minimal replicon of a cryptic Deinococcus
 radiodurans SARK plasmid and development of versatile Escherichia
 coli-D. radiodurans shuttle vectors
 JOURNAL Appl. Environ. Microbiol. 66 (9), 3856-3867 (2000)
 MEDLINE 20422197
 PUBMED 10966401
 REFERENCE 2 (bases 1 to 11910)
 AUTHORS Meima,R. and Lidstrom,M.E.
 TITLE Direct Submission
 JOURNAL Submitted (18-NOV-1999) Chemical Engineering, University of
 Washington, Benson Hall, PO Box 351750, Seattle, WA 98195, USA
 FEATURES
 Source
 1..11910
 /organism="Shuttle vector p13"
 /specific_host="Deinococcus radiodurans SARK"
 /db_xref="taxon:115340"
 /note="derived from pUE10"

repeat_region /note="L1 repeat: matches 4449. .5390 of consensus"
 40148. .41111
 /note="L1MD1 repeat: matches 1. .968 of consensus"
 complement(41972. .42157)
 repeat_region /note="MLTIF repeat: matches 541. .354 of consensus"
 complement(43606. .43671)
 /note="MIR2 repeat: matches 134. .71 of consensus"
 45479. .43585
 /note="MER45 repeat: matches 26. .133 of consensus"
 48242. .48312
 /note="MER45 repeat: matches 98. .176 of consensus"
 49604. .49826
 /note="AluB repeat: matches 82. .301 of consensus"
 49936. .30472
 /note="L1PA13 repeat: matches 41. .573 of consensus"
 complement(50932. .51134)
 /note="MER20 repeat: matches 218. .18 of consensus"
 51378. .51596
 /note="MER20 repeat: matches 1. .215 of consensus"
 complement(52293. .52552)
 repeat_region /note="MLTIC repeat: matches 466. .186 of consensus"
 complement(52555. .52908)
 /note="THB1B repeat: matches 362. .1 of consensus"
 complement(52914. .53071)
 /note="MLTIC repeat: matches 182. .8 of consensus"
 53212. .53621
 /note="WSTC repeat: matches 1. .403 of consensus"
 55202. .55295
 /note="MIR2 repeat: matches 1. .94 of consensus"
 55296. .55504
 /note="MER3 repeat: matches 1. .209 of consensus"
 complement(55783. .56684)
 /note="L1PA10 repeat: matches 911. .1 of consensus"
 complement(56534. .57144)
 /note="L1 repeat: matches 5390. .4785 of consensus"
 complement(57148. .59003)
 /note="L1 repeat: matches 4720. .2850 of consensus"
 59735. .59757
 61714. .61945
 /note="match: STS L24834"
 complement(63228. .63358)
 /note="MIR repeat: matches 247. .100 of consensus"
 complement(64335. .64455)
 /note="L1MB3 repeat: matches 909. .790 of consensus"
 complement(65455. .65744)
 /note="AluJo repeat: matches 289. .2 of consensus"
 66343. .66557
 /note="MIR repeat: matches 41. .261 of consensus"
 66912. .67656
 /note="L1MC3 repeat: matches 1660. .2409 of consensus"
 complement(68490. .68794)
 /note="AluJ repeat: matches 301. .1 of consensus"
 /note="AluJ repeat: matches 68873. .69190)
 complement(68873. .69190)
 /note="AluJ repeat: matches 298. .1 of consensus"
 69217. .69532
 /note="MER33 repeat: matches 1. .324 of consensus"
 complement(69638. .69703)
 /note="MIR repeat: matches 129. .64 of consensus"
 71210. .71340
 /note="MIR repeat: matches 83. .206 of consensus"
 72426. .72792
 /note="match: GSS A0053969"
 complement(72923. .72973)
 /note="MIR repeat: matches 139. .89 of consensus"
 complement(75007. .75196)
 /note="MER5A repeat: matches 189. .1 of consensus"
 complement(75386. .75425)
 /note="L1MB3 repeat: matches 406. .367 of consensus"
 75508. .75800
 /note="AluJo repeat: matches 1. .288 of consensus"
 complement(76744. .76840)
 /note="MLTIF repeat: matches 182. .91 of consensus"
 complement(78342. .78521)
 repeat_region

repeat_region /note="MIR repeat: matches 216. .35 of consensus"
 complement(78988. .79020)
 /note="MIR2 repeat: matches 144. .112 of consensus"
 complement(79137. .79186)
 repeat_region /note="MIR2 repeat: matches 134. .85 of consensus"
 complement(79151. .79393)
 /note="MIR repeat: matches 242. .3 of consensus"
 complement(79492. .79666)
 /note="MER5B repeat: matches 178. .5 of consensus"
 complement(79777. .79863)
 /note="MER5A repeat: matches 109. .27 of consensus"
 79894. .79970
 /note="MER5A repeat: matches 11. .91 of consensus"
 complement(80377. .80431)
 /note="MIR2 repeat: matches 141. .87 of consensus"
 complement(80389. .80462)
 /note="MIR repeat: matches 252. .177 of consensus"
 complement(80455. .80578)
 /note="MIR repeat: matches 147. .29 of consensus"
 repeat_region

Query Match 64.0%; Score 16; DB 92; Length 117968;
 Best Local Similarity 100.0%; Pred. No. 32;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 atgcgcagattccga 18
 |||
 Db 97872 ATGCCAGATTCCGA 97857

RESULT 27
 AC017911
 LOCUS
 DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***
 AC017911
 ACCESSION
 VERSION
 KEYWORDS HTG; HTGS; PHASE2.
 SOURCE
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 118100)
 AUTHORS Adams, M. and Venter, J.C.
 TITLE Direct Submission
 JOURNAL Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
 Rockville, MD, USA

COMMENT This sequence was identified as CDM:10212606 by the submitter.
 For more information on this record e-mail to fly@celera.com.
 * NOTE: This is a 'working draft' sequence.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.

FEATURES
 source
 1. .118100
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"

BASE COUNT 32824 a 26059 c 26550 g 32667 t
 ORIGIN

Query Match 64.0%; Score 16; DB 65; Length 118100;
 Best Local Similarity 100.0%; Pred. No. 32;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 gcatttcgattccgct 25
 |||
 Db 63425 GCATTTCGATTCGCT 63440

RESULT 28
 AP002817/c
 LOCUS AP002817 140825 bp DNA PLN 12-AUG-2000

DEFINITION	Oryza sativa genomic DNA, chromosome 1, PAC clone:P0699D11.	CDS	GSDDCLADAASYLGAELQRLPRCG"
ACCESSION	AP002817 BA000010		complement(join(15587..15727,15868..15915))
VERSION	AP002817.1 GI:9558510		/note="hypothetical protein"
KEYWORDS			/codon_start=1
SOURCE	Oryza sativa (cultivar:Nipponbare) DNA, clone:P0699D11.		/protein_id="BAB03431.1"
ORGANISM	Oryza sativa		/db_xref="GI:9558513"
	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.		/translation="MNNPHVYEAGRGKSHSTVYKGRKKKSIEYFAVKSVDKORSKVL
REFERENCE	1 (bases 1 to 140825)		NEVSAAPSSPCSFISSW"
AUTHORS	Sasaki,T., Matsumoto,T. and Yamamoto,K.		complement(join(18733..18801,19343..19489,19843..19905,
TITLE	Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC clone:P0699D11		20016..20087,20611..20691,20769..20819,20976..21072,
JOURNAL	Published Only in DataBase (2000) In press		21611..21753))
REFERENCE	2 (bases 1 to 140825)		/note="ESTS C72327(E1437),AU078752(E1437) correspond to a
AUTHORS	Sasaki,T., Matsumoto,T. and Yamamoto,K.		region of the predicted gene.
TITLE	Direct Submission		Similar to Archaeoglobus fulgidus competence-damage
JOURNAL	Submitted (26-JUL-2000) Takuji Sasaki, National Institute of Agrobiological Resources, Rice Genome Research Program; Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan		protein (AE000949)"
REFERENCE	(E-mail:tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/, Tel:81-298-38-7441, Fax:81-298-38-7468)		/codon_start=1
COMMENT	The orientation of the sequence is from sp6 to T7 of the PAC clone. Genes were predicted from the integrated results of the following:GENSCAN1.0, BLASTN2.0, BLASTX2.0 as well as SplicePredictor (October1998 version). The genomic sequence was searched against the non-redundant database NRP(PIR,SWISSPROT, GENPEPT, PDB) from MAFF DNABank and the cDNA sequence database at RGP. Protein similarities of the coding regions were searched against NRP with BLASTP2.0. ESTs represent the identified cDNA sequences using BLASTN2.0 with the corresponding DDBJ accession no. and RGP clone ID.		/protein_id="BAB03432.1"
	This sequence of P0699D11 clone has an overlap with P0462H08 clone, DDBJ:AP002525 at the 5' end. The sequence of this clone starts at the position 114,739 of P0462H08. This sequence of P0699D11 clone has an overlap with P0469E09 clone, DDBJ:AP001366 at the 3' end. The sequence of this clone ends at the position 59,396 of P0469E09. Detailed information on assemble quality together with annotation of this entry at		/db_xref="GI:9558514"
	http://rgp.dna.affrc.go.jp/GenomeSeq.html.		/translation="MISDGLERAGRTKKNISVSSNNTNSTEVEQTSRASIIVVG
FEATURES	Location/Qualifiers		DEILFTQVEDKLGAGLCKKLHAIGRWVSHVAVVNEIDSAVEVERCKSTDDMYFLVG
source	1..140825		GLGPLHSDISLAGAKAFQVRLAPDEFEYLSQIDGNYTGDRENMALLPEGITELL
	/organism="Oryza sativa"		HKMLPLPLIKCKNVVLAATNVDELETEWGLDQESGLVMAKSFVSKHLCTSLLD
	/cultivar="Nipponbare"		VKIAPVVAKLCIDFSDVYIGMI"
	/db_xref="taxon:4530"		join(30551..30609,30919..31139,31223..31578)
	/chromosome="1"		/note="Similar to Arabidopsis thaliana chromosome 4, BAC
	/clone="P0699D11"		clone F10W23; hypothetical protein (AL035440)"
	join(527..679,925..1114,2076..2467,3067..3555)		/codon_start=1
	/note="EST C22619(S11214) corresponds to a region of the predicted gene.		/protein_id="BAB03433.1"
	Similar to Arabidopsis thaliana chromosome 2, BAC clone T17D12; putative protein kinase (AC006587)"		/db_xref="GI:9558515"
	/codon_start=1		/translation="MUCGKRGRRREGGENEDDIVAKRSMGCPDPLHLCKSTISKVP
	/protein_id="BAB03429.1"		QLVIGIKVLEPNRYAGSLNCTKPVTRGSSVVLTPVLVYTAESWNPRIVRKRGHH
	/db_xref="GI:9558511"		HGLDROQGAGAEFDEGEVYTDGLLHQHQPVRIPSEAAATVRRRGWATOSSVGED
	/translation="MVGIRNKGHEATIFTLRELADATNMFSTECLLRGGRGVSKY		GDAAMVSPHKVVARAAHSHSVLEGAGMLKGCDDLHVNRNAVLCRTGFLD"
	AFLNDQVAVKQDLNGLGNREFLVEVLMLSLLHHPNLVKLFYCVGDGDORLLIYE		complement(join(31912..32455,32698..33684,34091..34341))
	YMPGSLSEDRLUHDLPQGEPLDWTTRNKIAADAAAGLEYLHDEAI PAVYIRKIPSN		/codon_start=1
	LLGEGYNKLSDFGLKGPVGDKTHVTVMGTHTGYCAPEYLSLTKLTKISDIYSFG		/protein_id="BAB03434.1"
	VVFLELITGRALDSNRPDEQDLVAMARPLFKDRKFFPKMADPSLGHFPRGLFQA		/db_xref="GI:9558516"
	LAIAMCLOEKAKNRPSIREVAVALSYSATQESQNTAARHTLPGPSVPRVLDNQIN		/translation="MRKKRTNLLIIVVGEATFFFFEERCRRRAGVGSEVERRGDSKP
	QDTSLPQSGVHMPPLAGTDHMQVEKNCSSSRHPRGRVTPNGADRERALEAENV		SRRGHPATARAAPLRREOKRAREDELATASGSGTPEGWPHHPLSPSPLLSLSH
	WVEAWRQEKTSKMR"		SRLFSRGLGEHGGGAESGASGSGGAAAGAAAGAAAGGCGWGGAAPTSP
	/note="ESTS AU093260(C52190),AU062991(C52190) correspond to a region of the predicted gene.		RHFALSSAAPGARDGAGLVGGVGGGTGEAGARSRRQIGOERHGMKAPSV
	hypothetical protein"		VCCCFPTVAELVVLAVVCPAALCHLVAVRGHRRVCSAKQKMGELLTLTDVTPRS
	/codon_start=1		LAATAAKARKVEVEFTATPNAELHGEAVRMKGLQALHWVARLLSAGEDARPA
	/protein_id="BAB03430.1"		DLRVLLPACLPSPVILPLPRHVSLSCLSPRLKSLTSLSPKRGNGDPSRRSR
	/db_xref="GI:9558512"		TEGNIKIKERENSTIAQKTCRSRSPATPPTVAVRPLTNHPIRQAPQRCR
	/translation="MDDFLEPTLELCRAFTIRIGTSSKI VALSKEPALLVDSAFMSI		EEGGGGGDAALPACLLAAYSLAASPVKVTEPTKPAVRPLTNHPIRQAPQRCR
	AVDQSCVMDICDFGNGMIFDLVGSDDPHISDLASDCLVLLKAPREATVGLLTN		SSSFSSSSLSRSGSKPARCCTVARERLFWRRVRQLRATSIHAEARAAHVIVVLV
	LPKLSVLDLLKHGTCRLRLTRLLYCLAFSCRQYLAQGMIVSISLSALMRVEALVSFAK		VKLPPAREACPRSTAPASAROGPPHRRR"
			join(35382..35472,36089..36172,36426..36530,37078..37337,
			38995..39114)
			/note="hypothetical protein"
			/codon_start=1
			/protein_id="BAB03435.1"
			/db_xref="GI:9558517"
			/translation="MQLDASDFHQOELTLKLRDGLKSLSSLESSGKSYDLVSFAAQOP
			DATVFLPSGPEAKAVGIARRVSLVYPDPEELEKPLGLSLTPSSRKSAGSKYKRSR
			FTSGKNRSQGSQAQSLGQSAFPAHKHOKRDESLSGSHNSVSGSKSAGSQVRGDS
			GTTSEVPTQPEVSKSKKKLDPRKRRTPYDRHKAQELDRATVAVDSSLYSPKPY
			Q"
			join(40731..40753,43157..43538,45516..45821,47347..47893,
			48657..48934,48983..50395)
			/note="Similar to Arabidopsis thaliana chromosome II BAC
			clone r27A16; hypothetical protein (AC005496)"
			/codon_start=1
			/protein_id="BAB03436.1"
			/db_xref="GI:9558518"
			/translation="MYSLRKIRPPPPALAAARDSPLRWPPPPGRLPSWIRWGRDR
			RRLIPYAMSRGPPGSGRGRENEEERSRRGGGEGGEGEEIKGMGRKEEREKE
			ETRLGAGRMRRDAGTGSFADFLSLFNKSLVSLWLTAKDEIKQLANRRRRPDIGIGG
			HGAAKDLPIDRSIYEKLAGHSGIHOLASCWHLGPCSSCKAGLWRRRSKASAL
			ILSSCDLNPIDRSIYEKLAGHSGIHOLASCWHLGPCSSCKAGLWRRRSKASAL
			HLSSCDLNPIDRSIYEKLAGHSGIHOLASCWHLGPCSSCKAGLWRRRSKASAL
			FSLNLGFRFRPPPTSIVRPTAMRTVRSIVTIPYVDGVRVWRKLIAPYTSIAISMSS

CDS

CDS

FEATURES

SLNPNMQTVRCIGSGALSGLGIGTPIHRAAKPIAPFGYCNFPFFFLKKYSKTEALAH
 AYTAHSSINRAHTPIRKIDILYKOARNLCTSPCKNRSHINISAOCHTNRLPAQ
 YRYDLSEARHPQREAREARETHADRTSSTENRPPPPATAERWPPPPPPPPPP
 DLHRAHAYLLPGGHHARVASALLASLRLPLROHAAALVRRVHPSVLSRAARLRG
 RGGGGLAAOLHSLVLAGHAAADPHASASIVQAYSCGSASAKRPPEDTAASADVVSW
 NVMIDGYVSGDLARARELFDVWPGNRVSVMTVIGAYAKOPPEAEIEFRMVGVEG
 IEPDGVALLSVLSACDGLGVLDGEMWHRVFLRGLEAVEFLRMKEKNVSPNDITFLAVLSVC
 ALUEFGEQKISIVTWTMIAGFALHGLGLEAVEFLRMKEKNVSPNDITFLAVLSVC
 SHVGLDGLRWTFKTMVSOYKIKPRVHEHGLGCMIDLLGRAGCLMEARGLIQDMPKANA
 ATWGLLAAARHGTDELCEQALLHLIELEPHNSGNYILLNIAEQRWDAVRELRI
 SMNRGLRNVPCASSIDVDMVHEFTSRDGSHPSLHKIREVLCALNSKINSVGHIALI
 PESLHDVEEG"

complement(join(52441..53547,53887..54014,54128..54165,
 54272..54487,54709..54870,55838..56016,56150..56256,
 56375..56539,56658..56741,56821..57040))
 /note="Similar to Arabidopsis thaliana chromosome I BAC
 clone T5A14; kinesin (AC005223)"
 /codon_start=1
 /protein_id="BAB03437.1"
 /db_xref="GI:9558519"

CDS

/translation="MDVQPARTMRNLPTLSLMGFNKLHLPSTWIESVSHIIDGLSPT
 KPMQVMVEKDEISDDNTESEAKYQKIDELVSLNAQLQITLQRRALNNYLDLKG
 NTRVFCRIPEHEESYSRNLTLDSENVFLKVAETKRKQYKFDKVDQFSTQGVF
 SEVERVKSALDGVNVCIFYAGOTSGTKTYMEGKPTNLGVIPIRGLOTLENOASECNN
 RELFTFMLEIYMGNIROLLAPRSKNGIKNVPSLKSIPDGGIEIEDLVAVTVNSF
 QEVKLYEMGTLSRTASTMANSTSSRSCHLIRISLTINATERRKATSKLUMIDLGG
 SERLVKATGRKLKEGKAINLSALGDIDALQTKRPHVYRNSKLTQVLRDLSGC
 ESKTLMVHIISPDGDLCTICTLGTGATRVRSIRLESEPEMKARKETLLIDLGQV
 NDLEHEDDIRIKNESEMEHTGPQPTVYSNFDMSLSSEELKATKEDLLIDGQV
 REASRLPRMPTASSQHRIGLNRTPIINRLKPPVPPRRPSSVVAESVMVPA
 APWQSECSSEKSMSTSMNTVPSIRDGTQSDASEVEIKQVIFSEHEKSHDOVTC
 YTDYPLASRDIOIKIEKGIVDIDNLHQOIVKSTPFSKWLVIDIPGVEAEIHSV
 SIPSPTWACIKEDSOVKDEVGLTLOSSTIYVEDIKQSKDNOFTAKELCTPPKFF
 SSNNEVKGHNEHPYHGRPRRSQLEENLTLEKPNMDSKSHRSHDDKKTGNVLSY
 TGE"

join(57603..57645,57742..57843,58195..58344,59425..59479,
 59537..59589,61304..61785)
 /note="hypothetical protein"
 /codon_start=1
 /protein_id="BAB03438.1"
 /db_xref="GI:9558520"

CDS

/translation="MKQDKGNMRCFLQGSFSAPIYHOHVTIKGEECTPTDTPSERSR
 SEYAKLMDIYIPWMAITFGNMCANENHSRAVIRKSGPTMCHAELESEATAIS
 LLASFAGAPFELSKRRSGDVLHMAISGVEAAALAPLPPOPLPPGDRMPAAGG
 LCLSPASARLSPPLPQRPQPMRDRAAAGLRLSPNARRREIKRPLPPTSTSP
 AIQPCRWPPPLPADPATGELELYSKRCRSCVERIPTVELLPMVGELELFSSTVKL
 LPSPCSPAMSLAPSSRA"

complement(62330..63121)
 /note="Similar to Arabidopsis thaliana DNA chromosome 4,
 BAC clone M3E9; hypothetical protein (AL022223)"
 /codon_start=1
 /protein_id="BAB03439.1"
 /db_xref="GI:9558521"

CDS

Query Match 64.0%; Score 16; DB 13; Length 140825;
 Best Local Similarity 100.0%; Pred. No. 31;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 2 catcgccagcatttcg 17
 |||||
 Db 89622 CATCGCCAGCATTTCG 89607

RESULT 29
 AP001366/c
 LOCUS AP001366 146081 bp DNA PLN 12-AUG-2000
 DEFINITION Oryza sativa genomic DNA, chromosome 1, PAC clone:P0469E09.
 ACCESSION AP001366 BA000010
 VERSION AP001366.1 GI:7228436
 KEYWORDS
 SOURCE Oryza sativa (cultivar:Nipponbare) DNA, clone:P0469E09.
 ORGANISM Oryza sativa
 Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL

Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae;
 Oryza.
 1 (bases 1 to 146081)
 Sasaki,T., Matsumoto,T. and Yamamoto,K.
 Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
 clone:P0469E09
 Published Only in DataBase (2000) In press
 2 (bases 1 to 146081)
 Sasaki,T., Matsumoto,T. and Yamamoto,K.
 Direct Submission
 Submitted (09-MAR-2000) to the DDBJ/EMBL/Genbank databases. Takuji
 Sasaki, National Institute of Agricultural Resources, Rice Genome
 Research Program; Kannonnai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
 (E-mail:tsasaki@agr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
 Tel:81-298-38-7441, Fax:81-298-38-7468)
 The orientation of the sequence is from SP6 to T7 of the PAC clone.
 Genes were predicted from the integrated results of the
 following:GENSCAN1.0, BLASTN2.0, BLASTX2.0 as well as
 SplicePredictor (October 1998 version). The genomic sequence was
 searched against the non-redundant database NRP (PIR, SWISSPROT,
 GENPEPT, PDB) from MAF DNA bank and the cDNA sequence database at
 RGP. Protein similarities of the coding regions were searched
 against NRP with BLASTP2.0. ESTs represent the identified cDNA
 sequences using BLASTN 2.0 with the corresponding DDBJ accession
 no. and RGP clone ID.
 Detailed information on overlap and assemble quality together with
 annotation of this entry at
 http://www.dna.affrc.go.jp:82/genomicdata/GenomeFinished.html.

FEATURES
 source

Location/Qualifiers
 1..146081
 /organism="Oryza sativa"
 /cultivar="Nipponbare"
 /db_xref="taxon:4530"
 /chromosome="1"
 /clone="P0469E09"
 join(5902..5980,6229..6272,8134..8295)
 /note="hypothetical protein"
 /codon_start=1
 /protein_id="BAA92397.1"
 /db_xref="GI:7228437"
 /translation="MCAVRRSSFLVVVVVFAILLTMAAGRKMLINKHQVQSMETS
 DDESHMQGQDDEMLAMVIERILRQVKINDYDTPPTMAKPAKEIPN"
 9673..9989
 /note="5' LTR"
 join(10528..12588,12673..13083,13175..13473,13550..13808,
 13901..14083,14241..14306,14743..15045,15216..15479)
 /note="EST C28952(C62945) corresponds to a region of the
 predicted gene.
 Similar to maize transposon MuDR mudra protein isolog
 (AC003981)"
 /codon_start=1
 /protein_id="BAA92398.1"
 /db_xref="GI:7228438"

CDS

LTR

CDS

VNPLGLPTWRPERECRGRCHVIRSVRHRTGQRCFCVCPNIVDDDFVFIHAWMEYER
 RVEIASLDRNNPLGLPKWSPKCGCRDRWQVATSFACQFRRRIYFVCPNVDNDFVV
 CLTIYKIRYDLHKITQTE"
 16155..16469
 /note="3' LTR"
 join(23418..23748,25558..25589,25715..26480,26838..26965)
 /note="hypothetical protein"
 /codon_start=1
 /protein_id="BAA92399.1"
 /db_xref="GI:7228439"
 /translation="MGANCIIAKERPOCVPIEVSAFNRNVRHSPSFSRWDNRTHI
 EDIMWPAFNSHSSGSIIRPETKSGSIAPDTGFSNGSPDMENKIKCHKSDRKRSS
 KIARSDLRSTNSSEAKLSRSLDTVSVASDSKWSISVPSPTPAISRADSSSS
 RGHSLPTDASMRKARSPGYQLYIQVSDSKIPISLRSNEGASPPGESSMLSVCSN
 DLSVSGHSGSDGSMRTFSMAVSSARWSVDSSELLGSVSKMTKSNANPNPTH
 SPQVECKLCKLLKRTWNAQELAVVALLCGHYHADCLDSLTADRYDPPCY
 CTHGEKTVKLFKLESTKTKPKNVIVDNLGSGSKHQEKIPLEBELYSQARSM
 RELKSRIGANDMTLLNDRLEGPS"
 complement(join(27953..28624,29565..29918,30369..30646,
 31026..31191))
 /note="ESTs C96615(C10106),C26336(C12127),D21959(C10106)
 correspond to a region of the predicted gene.
 Similar to NAM (AL021889)"
 /codon_start=1
 /protein_id="BAA92400.1"
 /db_xref="GI:7228440"
 /translation="MESLRDMVLPPGFPKPDTELISHYLLKKKHGQKIEVEIIEPV
 DIVKHPWDLPAKCDPTQDNKWHFAARDKRYPNGRSNRATVAGYWKSTGDKRAIK
 MGKOTIGTKTIVFHGPRPTGRTTIEHVEYIDEREQACPMKDAVLCRITKRN
 DWTGNGNELDSHPDPYDAPPSVISTEOLNPAEPVVGVEAAPVTVAEPPGVTT
 AITANIPSSDDILNDLWELNLEDFDFPEQSLASADLSFDEONVSSNVGALAPKE
 QDYSNENVVDDTEYLLPEDVINLHPGTDNMLNQLDQPIQYATDWSKATQKE
 ELWSPQNAEPQSNDAADNGIIRYRSMTKTPETSPQPKGTQAKMRGVINGMATSS
 SESINOTIKFENSGRIVEHQNAQHDVASTKRSACKPSTELSSNRFGLRGIRNAPAG
 CSDARNMILVAGFAIGVAVVALHIGQRLGLSQDQHT"
 complement(38747..38998)
 /note="hypothetical protein"
 /codon_start=1
 /protein_id="BAA92401.1"
 /db_xref="GI:7228441"
 /translation="MMGHVLRADTARTAPDNMNAVPGPEVRHVRHGTPEVIVLCRPD
 DRGCRASILMAMRLGLELVHFGCCYDDPEPHGPSCTSI"
 complement(39315..39726)
 /note="3' LTR"
 complement(join(40209..40396,40917..43545))
 /note="EST C28952(C62945) corresponds to a region of the
 predicted gene.
 Similar to maize transposon MuDR mudra protein isolog.
 (AC003981)"
 /codon_start=1
 /protein_id="BAA92402.1"
 /db_xref="GI:7228442"
 /translation="MSKVTFQIVHGEGRNIRFGPDGVDLSDFVMTSGIDRPAERTFQ
 SIYSWLLRGRFIDQEVYTVMSVVSVRATGEYFELWMLPMDSTDAWRVYEMAFEWSWP
 LTVFSVQERDINVSQOTDEGPIINAGDVVGPSSMONENQPREOAGMADGERGVG
 IIVDEMERSDNEVDDDDASSDEEGDVMAFDHANEDESLGVISEGDHVPWEYKENEV
 IEGARYAHKDEMEAVHAWVASIQREFRVVKSNIYVEVRCMEKEDCPWRVHAYKGVN
 DTKVSVITAEKYLQCYGKERYHRNITSAPMASEMISVVGNIQFEPKSIIRHLENKFK
 YTISYAKAWRAKQKIIEMRYGTFEASYNLPRLLIAIARNNNTYYDLHTFTSVDORT
 KSVLQRAFSLGACINAFVHCRPVLICIDGTFMGKYRQGIILTAIGCNGNVLPMALW
 FVSENTFESYWFLEHVIAVVRMRPNVLIHHRHAGMLRAIDYLONGWDEKGLPAKW
 DVSRCMRHMGANFYKPKNHILMELFKLCAQNOCKKFNELWDKLDELTTKOTDE
 QSRPLVEGDEPPIPLGALHDDPTMRRSGSSIRNPTQWIENTEPKEKSLLEFDGGS
 RYGLMTNLVNWVRGVLPVLAIVELFHGTOYFRDRFKIGPSMADNIVF
 GNVTYKYMEDIKKARHVRVQGTQVHRYEIMCVSRREGIYKRVQEEYCVLKADGG
 CTCSCMKPKLHLPCSHVLAAGDCGIPPNVYSNYPERKAIPTWSEETLYGFGISG
 YTTLSAOFVYLPDPSKLRVKKRRGRTIRINDMDESEAGGRTLCRSCDLRGHTYKRC
 PKNAEVPSGADSPSGOASDVLEFTHWVLEQLMTLTHFRCTTVVSDPVRTVHNSHTE
 KRQLELSLSTFGAAYLHKVHWAH"
 complement(43999..44385)
 /note="5' LTR"
 join(44572..44723,44846..44938,46366..46378)
 /note="hypothetical protein"

/codon_start=1
 /protein_id="BAA92403.1"
 /db_xref="GI:7228443"
 /translation="MALPRGVRSPKPMAAAAASASPPQPRGHMLVLTGTFGVGRYYSQ
 RLIAQGRVSGTCTSPAKKTELEMLGMDASTFDPATSRSCVM"
 complement(join(48119..48257,48429..48703,49975..50139))
 /note="hypothetical protein"
 /codon_start=1
 /protein_id="BAA92404.1"
 /db_xref="GI:7228444"
 /translation="MALVNDKDEDDVNAPACARRAAPLVIGAPWGFFVLQARLGEBC
 EPRFRARKWAILISFCALSLFLSKNTRIYVLCLYLRKNYRPLIVIKESRRTKCCRIT
 ASIARHGDELVLCLARLASSALDRVFKILKTSDFVLKMKVICINVPVHISRTH
 GSSMYSTLPDAPVYRSTERLAGGIVICTTIGG"
 complement(join(51918..52637,52684..53346))
 /note="Similar to Arabidopsis thaliana chromosome 1 YAC
 YUPH12R sequence, C2-HC type zinc finger protein
 C.e-Mytl. (AC002986)"
 /codon_start=1
 /protein_id="BAA92405.1"
 /db_xref="GI:7228445"
 /translation="MASCINGELHVRVHRLVKASDEIRPHVLAVSNLILPRTI
 QVSMFCIYKPKSTGGDFHVVAAFAAGLPSLLNHHFFPLAGRIVSNPCSGLPETIHCNQ
 CAELVVGADVALASLDYGTGVSGVKILLPYAGDVALSVQVVSFACGGFTVAGTTH
 VVVDGSAISMLVSAWSLARSCTLAAGARNPHDRSVFRSPSPSGASLDFAFTPLDG
 ARQADJARLREQARTRVOAVSAYLWKALAAVVGSRDARCMVMWVMDGRRRLTSSS
 PELNRAKRSYGNVTTFVATMEETQKPLAEVASMARDIAAPAGGEHFOELVOW
 VEEHAGKORTYIDTASVSGSPASVTAFAFEVDTRGHGHAAMALPTSSSARLCT
 GFVQTAARPPGGDGSWIASALLPRLAAALSDERLGRIFRVPVTAEYLGRLPASGSSA
 KRAGMITSKI"
 complement(57570..58009)
 /note="3' LTR"
 complement(join(58018..58891,58980..62098))
 /note="Similar to Arabidopsis thaliana chromosome II BAC
 F9013 genomic sequence, putative retroelement pol
 polyprotein. (AC006248)"
 /codon_start=1
 /protein_id="BAA92406.1"
 /db_xref="GI:7228446"
 /translation="MRVNLQAGLWTAIDPGVAFEDFMAALSAILOAVPREMLRGLA
 KHDTHAGDAIKTWVRGVDRVREAKEQFRQFESMRFKSPETPEEFAMRLTAVVAD
 IHDMGVMEDEHVNKKLVRVPKPKYVAISLEQLDLVKTHALEELVGRLSYVDSYSD
 DEEGSDGKLYLTREQVQARVQREKSGNKGARGPGTGNHARGPPGSPGKGA
 ATGANSRDISRKCFNCGDFGHYARQCRKPRQRREANLVQAAEEPTLLMAHVF
 VSLAGEATLGRTPGGOEYHLTKKVLIDHEDGEGEEVTRDFWFLDTPKRNHTGVRSAL
 AELDGVGVGTGVKFGDGSYIEIQGRGVTVFRCKNGDHRSLDTVYIIPKRLNLIISVRL

Query Match 64.0%; Score 16; DB 13; Length 146081;
 Best Local Similarity 100.0%; Pred. No. 31;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 catcgccagcatttcg 17
 |||||
 Db 8193 CATCGCCAGCAGCTTCG 8178
 RESULT 30
 AC007476/c
 LOCUS
 DEFINITION
 AC007476
 AC007476.5 GI:13374640
 HTG.
 fruit fly.
 Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 171705)
 Celniker,S.E., Adams,M.D., Krommiller,B., Tyler,D., Wan,K.H.,
 Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Branden,R.C.,
 Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,

Carlson, J.W., Center, A., Champe, M., Davenport, L.B., Dietz, S.M., Dodson, K., Dorsett, V., Doup, L.E., Doyle, C., Dresnek, D., Farfan, D., Ferreira, S., Frise, E., Galle, R.F., Garg, N.S., George, R.A., Gonzalez, M., Houck, J., Hoskins, R.A., Hostin, D., Howland, T.J., Ibegwam, C., Jalali, M., Kruse, D., Li, P., Mattei, B., Moshrefi, A., McIntosh, T.C., Moy, M., Murphy, B., Nelson, C., Nelson, K.A., Nunoo, J., Pacleb, J., Paragas, V., Park, S., Patel, S., Pfeiffer, B., Phouanavong, S., Pittman, G.S., Puri, V., Richards, B., Scapleton, M., Strong, R., Svirska, R., Tector, C., Williams, S.M., Zaveri, J.S., Smith, H.O., Rubin, G.M. and Venter, J.C.

Sequencing of Drosophila chromosome 2R, region 49A-49B
Unpublished
2 (bases 1 to 171705)

Celniker, S.E., Aghayani, A., Arcaina, T.T., Baxter, E., Blazef, R.G., Butenhorff, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L., Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L., Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L., Kim, E., Lee, B., Lewis, S., Li, P., Lomotan, M.A., Mazda, P., Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S., Pfeiffer, B., Poon, L., Sequeira, A., Sethi, H., Snir, E., Svirska, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and Rubin, G.M.

Direct Submission
Submitted (05-MAY-1999) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Mar 17, 2001 this sequence version replaced gi:5670614.
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720

This sequence was assembled using end sequences from a whole genome shotgun and from subclones of this BAC and its neighboring clones. For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive Web site (<http://www.fruitfly.org/sequence/>) or send email to bdg@fruitfly.berkeley.edu.

FEATURES

source
1. 171705
/organism="Drosophila melanogaster"
/strain="Y: cn bw sp"
/db_xref="taxon:7227"
/chromosome="2R"
/map="49A-49B"
/clone="BACR37P06 (D594)"
/clone.lib="RPCI-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial EcoRI in
pBACE3.6)"
47654 a 38437 c 38157 g 47457 t

BASE COUNT

ORIGIN
Query Match 64.0%; Score 16; DB 4; Length 171705;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 gcatttcgatctcgt 25
|||||

Db 106124 GCATTTCGATCTCCT 106109

RESULT 31

AC037435 AC037435 183937 bp DNA HTG 15-JAN-2001
LOCUS Mus musculus chromosome 11 clone RP23-98I3 map 11, WORKING DRAFT
DEFINITION SEQUENCE, 24 unordered pieces.
AC037435
AC037435
AC037435.2 GI:12229535
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1. (bases 1 to 183937)

AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F., Boguslavsky, L., Bonkhgalter, B., Brown, A., Burkett, G., Campione, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S., Dodge, S., Domino, M., Doyle, M., Ferrel, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Larocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., Locke, K., Macdonald, P., Marquis, N., McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Meldrum, J., Meneus, L., Mihova, T., Miranda, C., Mlenka, V., Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Olivari, T.M., Oliver, J., Peterson, K., Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission
Submitted (08-APR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jan 15, 2001 this sequence version replaced gi:7528163.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

TITLE
JOURNAL
COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WBIR
Web site: <http://www.seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L8882
Center clone name: 98_1_3
----- Summary Statistics
Sequencing vector: M13; M7815; 4% of reads
Sequencing method: Plasmid; n/a; 96% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 170056 bases at least Q40
Consensus quality: 177644 bases at least Q30
Consensus quality: 180041 bases at least Q20
Insert size: 188000; agarose-fp
Quality coverage: 4.7 in Q20 bases; agarose-fp
Quality coverage: 4.9 in Q20 base.
NOTE: This is a 'working draft' sequence. It currently
* consists of 24 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.

1 23881: contig of 23881 bp in length
* 23882 23981: gap of 100 bp
* 23982 24469: contig of 488 bp in length
* 24470 24569: gap of 100 bp
* 24570 24755: contig of 186 bp in length
* 24756 24855: gap of 100 bp
* 24856 25997: contig of 1142 bp in length
* 25998 26097: gap of 100 bp
* 26098 26764: contig of 667 bp in length
* 26765 26864: gap of 100 bp
* 26865 28426: contig of 1562 bp in length
* 28427 28526: gap of 100 bp
* 28527 30488: contig of 1962 bp in length

```

* 30489 30588: gap of 100 bp
* 30589 32514: contig of 1926 bp in length
* 32515 32614: gap of 100 bp
* 32615 36175: contig of 3561 bp in length
* 36176 36275: gap of 100 bp
* 36276 38734: contig of 2459 bp in length
* 38735 38834: gap of 100 bp
* 38835 40927: contig of 2093 bp in length
* 40928 41027: gap of 100 bp
* 41028 43453: contig of 2426 bp in length
* 43454 43553: gap of 100 bp
* 43554 46744: contig of 3191 bp in length
* 46745 46844: gap of 100 bp
* 46845 50369: contig of 3525 bp in length
* 50370 50469: gap of 100 bp
* 50470 56799: contig of 6330 bp in length
* 56800 56899: gap of 100 bp
* 56900 57617: contig of 718 bp in length
* 57618 57717: gap of 100 bp
* 57718 68517: contig of 10800 bp in length
* 68518 68617: gap of 100 bp
* 68618 82793: contig of 14176 bp in length
* 82794 82893: gap of 100 bp
* 82894 96371: contig of 13478 bp in length
* 96372 96471: gap of 100 bp
* 96472 112208: contig of 15737 bp in length
* 112209 112308: gap of 100 bp
* 112309 127004: contig of 14696 bp in length
* 127005 127104: gap of 100 bp
* 127105 147474: contig of 20370 bp in length
* 147475 147574: gap of 100 bp
* 147575 174126: contig of 26552 bp in length
* 174127 174226: gap of 100 bp
* 174227 183937: contig of 9711 bp in length.
FEATURES
    source
        Location/Qualifiers
            1..183937
                /organism="Mus musculus"
                /db_xref="taxon:10090"
                /chromosome="11"
                /map="11"
                /clone_lib="RPCI-23 Female Mouse BAC"
            1..23881
                /note="assembly_fragment"
                clone_end:SP6
                vector_side:left
            23982..24469
                /note="assembly_fragment"
            24570..24755
                /note="assembly_fragment"
            24856..25997
                /note="assembly_fragment"
            26098..26764
                /note="assembly_fragment"
            26865..28426
                /note="assembly_fragment"
            28527..30488
                /note="assembly_fragment"
            30589..32514
                /note="assembly_fragment"
            32615..36175
                /note="assembly_fragment"
            36276..38734
                /note="assembly_fragment"
            38835..40927
                /note="assembly_fragment"
            41028..43453
                /note="assembly_fragment"
            43554..46744
                /note="assembly_fragment"
            46845..50369
                /note="assembly_fragment"
            50470..56799

```

```

/note="assembly_fragment"
56900..57617
/note="assembly_fragment"
57718..68517
/note="assembly_fragment"
68618..82793
/note="assembly_fragment"
82894..96371
/note="assembly_fragment"
96472..112208
/note="assembly_fragment"
112309..127004
/note="assembly_fragment"
127105..147474
/note="assembly_fragment"
147575..174126
/note="assembly_fragment"
174227..183937
/note="assembly_fragment"
clone_end:T7
vector_side:right"
BASE COUNT 47977 a 43626 c 42743 g 47258 t 2333 others
ORIGIN

Query Match 64.0% Score 16; DB 71; Length 183937;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gcatccgcagcatttc 16
|||||
Db 90966 GCATGCCAGCAGCATTTC 90981

RESULT 32
AE003822/c
LOCUS
DEFINITION
Drosophila melanogaster genomic scaffold 142000013386047 section 21
of 52 complete sequence.
ACCESSION AE003822 AE002767
VERSION AE003822.2 GI:10727606
KEYWORDS HTG.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 266133)
AUTHORS
Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,
Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J.,
Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A.,
Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkov,D.,
Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferrieria,S.,
Fleischmann,W., Fosler,C., Gabrielian,A.E., Garg,N.S.,
Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
Guan,P., Harris,M., Harris,N.D., Harvey,D., Heiman,T.J.,
Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
Li,J., Li.Z., Liang,Y., Lin,X., Liu,X., Liu,X., McInosh,T.C.,
McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,

```

Morris, J., Moshrefi, A., Mount, S.M., Moy, M., Murphy, B., Murphy, L., Muzby, D.M., Nelson, D.L., Nelson, D.R., Nelson, K.A., Nixon, K., Nusskern, D.R., Pacleb, J.M., Palazzolo, M., Pittman, G.S., Pan, S., Pollard, J., Puri, V., Reese, M.G., Reinert, K., Remington, K., Saunders, R.D., Scheeler, F., Shen, H., Shue, B.C., Siden-Kiamos, I., Simpson, M., Skupski, M.P., Smith, T., Spier, E., Spradling, A.C., Stapleton, M., Strong, R., Sun, E., Svirska, R., Tector, C., Turner, R., Venter, E., Wang, A.H., Wang, X., Wang, Z.Y., Wasserman, D.A., Weinstein, G.M., Weissbach, J., Williams, S.M., Woodage, T., Worley, K.C., Wu, D., Yang, S., Yao, Q.A., Ye, J., Yeh, R.F., Zaveri, J.S., Zhan, M., Zhang, G., Zhao, Q., Zheng, L., Zheng, X.H., Zhong, F.N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H.O., Gibbs, R.A., Myers, E.W., Rubin, G.M. and Venter, J.C.

The genome sequence of *Drosophila melanogaster*
 Science 287 (5461), 2185-2195 (2000)
 20196006

TITLE
 JOURNAL
 MEDLINE
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT
 FEATURES
 Source

2 (bases 1 to 266133)
 Adams, M.D., Celniker, S.E., Gibbs, R.A., Rubin, G.M. and Venter, C.J.
 Direct Submission
 Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
 Rockville, MD, USA
 On Oct 9, 2000 this sequence version replaced gi:7303437.
 Location/Qualifiers
 1. 266133
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /chromosome="2R"
 join(392..425,1007..1155,5710..6105)
 /gene="Cam"
 /product="CT41657"
 /db_xref="FLYBASE:FBan0008472"
 /db_xref="FLYBASE:FBgn0000253"
 <392..>15192
 /gene="Cam"
 /note="CG8472"
 /map="48F5-48F5"
 /db_xref="FLYBASE:FBan0008472"
 /db_xref="FLYBASE:FBgn0000253"
 join(405..425,1007..1155,5710..5884,10160..10402,
 14732..15192)
 /gene="Cam"
 /product="CT41655"
 /db_xref="FLYBASE:FBan0008472"
 /db_xref="FLYBASE:FBgn0000253"
 join(596..619,1004..1155,5710..5884,10160..10402,
 14732..14876)
 /gene="Cam"
 /product="CT41659"
 /db_xref="FLYBASE:FBan0008472"
 /db_xref="FLYBASE:FBgn0000253"
 join(1153..1155,5710..5922)
 /gene="Cam"
 /note="Cam gene product [alt 1]"
 /codon_start=1
 /db_xref="FLYBASE:FBan0008472"
 /db_xref="FLYBASE:FBgn0000253"
 /protein_id="AAF58544.1"
 /db_xref="GI:7303486"
 /translation="MADQLTEQIAEFKEAFSLFDKDGDTITTKELGTVMRSLGNP
 TEALQDMINEVDADNGTIDFPEFLTMARKMKDTSDEEIREAFRVFDKDGNGFIS
 AAELRHVMNLGKLTDEVDIMREADIDGQGVNYEEFVTMTSK"
 join(1153..1155,5710..5884,10160..10402,14732..14760)
 /gene="Cam"
 /note="Cam gene product [alt 4]"
 /codon_start=1
 /db_xref="FLYBASE:FBan0008472"
 /db_xref="FLYBASE:FBgn0000253"
 /protein_id="AAF58542.1"
 /db_xref="GI:7303486"
 /translation="MADQLTEQIAEFKEAFSLFDKDGDTITTKELGTVMRSLGNP
 TEALQDMINEVDADNGTIDFPEFLTMARKMKDTSDEEIREAFRVFDKDGNGFIS
 AAELRHVMNLGKLTDEVDIMREADIDGQGVNYEEFVTMTSK"
 join(1153..1155,5710..5884,10160..10402,14732..14760)
 /gene="Cam"

/note="Cam gene product [alt 2]"
 /codon_start=1
 /db_xref="FLYBASE:FBan0008472"
 /db_xref="FLYBASE:FBgn0000253"
 /protein_id="AAF58541.1"
 /db_xref="GI:7303485"
 /translation="MADQLTEQIAEFKEAFSLFDKDGDTITTKELGTVMRSLGNP
 TEALQDMINEVDADNGTIDFPEFLTMARKMKDTSDEEIREAFRVFDKDGNGFIS
 AAELRHVMNLGKLTDEVDIMREADIDGQGVNYEEFVTMTSK"
 join(5527..5642,5710..5884,10160..10402,14732..15192)
 /gene="Cam"
 /product="CT24787"
 /db_xref="FLYBASE:FBan0008472"
 /db_xref="FLYBASE:FBgn0000253"
 join(5640..5642,5710..5884,10160..10402,14732..14760)
 /gene="Cam"
 /note="Cam gene product [alt 3]"
 /codon_start=1
 /db_xref="FLYBASE:FBan0008472"
 /db_xref="FLYBASE:FBgn0000253"
 /protein_id="AAF58543.1"
 /db_xref="GI:7303487"
 /translation="MADQLTEQIAEFKEAFSLFDKDGDTITTKELGTVMRSLGNP
 TEALQDMINEVDADNGTIDFPEFLTMARKMKDTSDEEIREAFRVFDKDGNGFIS
 AAELRHVMNLGKLTDEVDIMREADIDGQGVNYEEFVTMTSK"
 join(<21254..21464,21584..>21600)
 /gene="CG13166"
 /product="CT32407"
 /db_xref="FLYBASE:FBan0013166"
 /db_xref="FLYBASE:FBgn0040758"
 /evidence=not_experimental
 <21254..>21600
 /gene="CG13166"
 /map="48F5-48F6"
 /db_xref="FLYBASE:FBan0013166"
 /db_xref="FLYBASE:FBgn0040758"
 /evidence=not_experimental
 join(21254..21464,21584..21600)
 /gene="CG13166"
 /note="CG13166 gene product"
 /codon_start=1
 /db_xref="FLYBASE:FBan0013166"
 /db_xref="FLYBASE:FBgn0040758"
 /evidence=not_experimental
 /protein_id="AAF58540.1"
 /db_xref="GI:7303484"
 /translation="MSRSGSGAVTGRHARVSTPTGRGSKTAYTPCRSEPPQDLAL
 AVPPPRNCIAIPMSSLOTANGPGTKELHNP"
 join(<34383..34566,36219..36635,37304..37431,37529..37772,
 37837..37951,38022..38163,38232..38305,38481..>38679)
 /gene="CG13165"
 /product="CT32406"
 /db_xref="FLYBASE:FBan0013165"
 /db_xref="FLYBASE:FBgn0033707"
 /evidence=not_experimental
 <34383..>38679
 /gene="CG13165"
 /map="48F5-48F8"
 /db_xref="FLYBASE:FBan0013165"
 /db_xref="FLYBASE:FBgn0033707"
 /evidence=not_experimental
 join(34383..34566,36219..36635,37304..37431,37529..37772,
 37837..37951,38022..38163,38232..38305,38481..38679)
 /gene="CG13165"
 /note="CG13165 gene product"
 /codon_start=1
 /db_xref="FLYBASE:FBan0013165"
 /db_xref="FLYBASE:FBgn0033707"
 /evidence=not_experimental
 /protein_id="AAF58539.1"
 /db_xref="GI:7303483"
 /translation="MFRKRFEGASKSGGGGKLGKQSKANRTKSRDMLGVQPPREIHY
 RTHLFFSPNRPDYDVGERCALSIPSTPLSSTAPSLALPNLPFLDMPQPLVDRR

```

/organism="Nicotiana glauca"
/strain="Samsun"
/db_xref="taxon:4097"
/dev_stage="mature pollen"
/clone_lib="lambda-EMBL3"
/clone="G10"
<1..577
/number=1
prim_transcript
join(1..577,676..976,1074..1389)
CDS
/codon_start=1
/product="pectate lyase"
/protein_id="CAA43414.1"
/db_xref="GI:119982"
/db_xref="SWISS-PROT:P40972"
/translation="MDVIRIVSFLLVLTFAALTATNIPRRLSNKKYKGPQRA
ENAIKQWRCDPNWNRQMRQADALGFGSLNAGLGRYIVVYTDNSDDVDVDPKPT
LRYGVIOKEPLWIFQKNNKIKLSRELIIVTSNKTIDGRGFNIHQNGAGIKIQASANI

```

```

/db_xref="taxon:4097"
/germline
/dev_stage="mature pollen"
/clone_lib="lambda EMBL3"
/clone="G10"
prim_transcript 108..1713
exon 108..501
number=1
join(108..501,1000..1300,1398..1713)
transit_peptide 325..399
join(325..399,1000..1300,1398..1713)
CDS
/codon_start=1
/product="pectate lyase"
/protein_id="CAA47630.1"
/db_xref="GI:19908"
/db_xref="SWISS-PROT:P40972"
/translation="MDVYRIRISVFLLVLTFAALTATNIPRQLSNKKYKPCRA"

```

ENAIKWCRCNDPNAENRQKMACALGFGSNAIGKGLRIYVVTNSDDVDVDPKPGT
 LRYGVIOKEPLWIFGKNMKIKLSRELIVTSNKTIDGRFNHVIQWAGIKIQASNI
 IISNIRTHNIVPTPGGLRSEDEHVLGRSGDEGDISIFSSHDIWIDHISMSRATDGL
 IDAVAATNITISNCHFTDHEKVMFGANDHYVLDKDKMTLAYNHFGKRLQRMPC
 RFGFFHLVNDYTHWERYAIGGSSGATIIISQGNRFIAEDELIVKEVYREKLTASVAE
 WKWWTWISDGDMMENATFTPSGDQNLIDKIDHLNLIKPEPSSKVGILTFSGALSCV
 KGRPC"

mat_peptide join(400..901,1000..1300,1398..1710)

intron /product="pectate lyase"

exon 902..999

exon /number=1

intron /number=2

exon 1301..1397

exon /number=2

exon 1398..1713

exon /number=3

BASE COUNT 546 a 338 c 353 g 476 t

ORIGIN

Query Match 60.0%; Score 15; DB 14; Length 1713;
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 qcatgccagcattt 15

|||||

Db 486 GCATGCCAGCATTT 472

RESULT 35

HSSYNTR05/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

ORIGIN

Query Match 60.0%; Score 15; DB 93; Length 1907;
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 catcgccagcatttc 16

|||||

Db 570 CATCGCCAGCATTTTC 556

RESULT 36

AB048915/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

AB048915 2616 bp mRNA PRI 20-SEP-2000

Macaca fascicularis brain cDNA, clone:QnpA-17839.

AB048915.1 GI:10241989

Macaca fascicularis adult male brain parietal lobe cDNA to mRNA,

clone_lib:macaque brain cDNA library QnpA clone:QnpA-17839.

Macaca fascicularis

Macaca fascicularis

Macaca fascicularis

Macaca fascicularis

Macaca fascicularis

Macaca fascicularis

Macaca fascicularis

Macaca fascicularis

Macaca fascicularis

Macaca fascicularis

Macaca fascicularis

Macaca fascicularis

Macaca fascicularis

Macaca fascicularis

Macaca fascicularis

Macaca fascicularis

Macaca fascicularis

Macaca fascicularis

Macaca fascicularis

Macaca fascicularis

Macaca fascicularis

Macaca fascicularis

Macaca fascicularis

Macaca fascicularis

Macaca fascicularis

Macaca fascicularis

Macaca fascicularis

Macaca fascicularis

Macaca fascicularis

Macaca fascicularis

Macaca fascicularis

Macaca fascicularis

Macaca fascicularis

Macaca fascicularis

Macaca fascicularis

Macaca fascicularis

Macaca fascicularis

Macaca fascicularis

Macaca fascicularis

Macaca fascicularis

Macaca fascicularis

Macaca fascicularis

Macaca fascicularis

Macaca fascicularis

Macaca fascicularis

Macaca fascicularis

Macaca fascicularis

Macaca fascicularis

Macaca fascicularis

Macaca fascicularis

Macaca fascicularis

Macaca fascicularis

Macaca fascicularis

Macaca fascicularis

Macaca fascicularis

Macaca fascicularis

Macaca fascicularis

Macaca fascicularis

Macaca fascicularis

Macaca fascicularis

Macaca fascicularis

Macaca fascicularis

Macaca fascicularis

Macaca fascicularis

Macaca fascicularis

Macaca fascicularis

Macaca fascicularis

Macaca fascicularis

Macaca fascicularis

Macaca fascicularis

```

Db 1743 CATGCCAGCATTT 1729
|||||
PSEIAAMH 3677 bp DNA BCT 26-APR-1993
P. syringae tryptophan 2-monoxygenase (iaam) and indoleacetamide
hydrolase (iaah) genes, complete cds.
M1035 M35690
M1035.1 GI:151289
indoleacetamide hydrolase; indoleacetic acid; tryptophan
2-monoxygenase; tryptophan-2-oxyreductase.
P. syringae savastanoi (strain EW2009) DNA, clones pLUC2, pCP3, and
pCJP12.
ORGANISM Pseudomonas syringae
Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
Pseudomonas.
REFERENCE 1 (bases 202 to 3677)
AUTHORS Yamada,T., Palm,C.J., Brooks,B. and Kosuge,T.
TITLE Nucleotide sequences of the Pseudomonas savastanoi indoleacetic
acid genes show homology with agrobacterium tumefaciens T-DNA
Proc. Natl. Acad. Sci. U.S.A. 82, 6522-6526 (1985)
JOURNAL 2 (bases 1 to 312)
AUTHORS Gaffney,T.D., da Costa e Silva,O., Yamada,T. and Kosuge,T.
TITLE Indoleacetic acid operon of Pseudomonas syringae subsp. savastanoi:
Transcription analysis and promoter identification
J. Bacteriol. 172, 5593-5601 (1990)
MEDLINE 91008926
COMMENT The iaah and iaam genes are part of an operon that is borne on a
plasmid, pIAA, in oleander strains of the pathogen. There appears
to be no genetic transformation of host tissue by P. savastanoi. No
Pribnow boxes were found, though potential ribosome binding sites
are located at positions 435-439 and 2164-2168. Expression of iaah
depends upon the promoter for iaam.
Draft entry and computer-readable sequence for [1] kindly submitted
by T. Yamada, 14-JAN-1986.
Draft entry and computer-readable sequence for [J. Bacteriol.
(1990) In press] kindly submitted
by T.D. Gaffney, 28-JUN-1990.
FEATURES
Source Location/Qualifiers
misc_feature 1..3677
/organism="Pseudomonas syringae"
/sub_species="savastanoi"
/db_xref="taxon:317"
19..47
/note="indoleacetic acid operon promoter"
mRNA 54..3537
/note="indoleacetic acid mRNA"
CDS 82..312
/note="promoter proximal ORF"
/codon_start=1
/transl_table=11
/protein_id="AAA25851.1"
/db_xref="GI:151290"
/translation="NPGRTPANWLDKAMVAVARIRQKPKQAVPAALLQAPHAMEFVW
STRTKFTTECHCPAYTPYTQASAWSASAVVSAY"
454..2127
/note="tryptophan 2-monoxygenase (EC 1.13.12.3)"
/codon_start=1
/transl_table=11
/protein_id="AAA25852.1"
/db_xref="GI:151291"
/translation="MYDHPNPSIDILVDYGFLLKCKEWTGGISYSGTPTPRVAIV
GAGISGLVATELLRAGVKVDVLYESRDVIRGVRVSVQVDFQTRPRYIAEMGAMRPPPS
ATGLPHYLKKGISTSTTPDPGVVDTELHRYHWPAGKKPPELFRVYEGWQSL
LSGVLLEGSLVAPUDITAMLSKGRLEEAIAWQWLNVRDCSFYNAIVGIFTRGH
PFGDWRAPDEPFEELGSGIGSGGFLPVFOAGFTELLRVINGYQSDRLPDCISS
LAARLADQSFQKALDRVCFRVRGRISREAEKIIQIOTAGEQRVDFRIVTSSNRAM
QMTCLTDSFSLRSVARAVRETHLTGSKSLFILTFRKFNKLPTTIQSDGLVRG
VYGLDYPDEPEGHGVLVLSYTWEDDAQKMLAMPDKKTRCQVLDLALAIHPFTFASYL
LPVDGYERYVLHDMWLTDPHSAGAFKLNYPGEDVYSQRLFFQPMPTANSPNKDITGLYL
AGCSCSFAGWGIVBQVATALNSACAVLIRSTGTGQLSKGNPLDCINASYR"
CDS 2170..3537
/note="indoleacetamide hydrolase"
/codon_start=1
/transl_table=11
/protein_id="AAA25853.1"
/db_xref="GI:151292"
/translation="MHEITITLESICOALADGEIAAAELRERALDTEARLARCNFIRE
GDVAVSGFGEADHAMKGTPLGMMPVSFKDNICVRGILPTAGTRGMSGFVSDDAAIVSQ
LRALGAVAGKNNHMHSELSFQVTSINPHWCTVGNVPVAGYCGAGSSGSAVAASGIVP
LSVDTGDSQSIPIPAFCIGTFRPTTGWSTAGLIPVSHTKDCVCLLRTAGDAGFL
YGLLSGKGSFPLSPACRIGLPSVMSDDLDEVEVERACVNSALSLRKTGTFEIDD
ADIVELNLTITFTVPLYEFFADLAQLSLSLGKGHGIIHIFIAQVDDANVKGIIHHHLE
GALKPAHYLSLQNGELLKRMDELFAHNIIELLGYPTVPCRPVHLDRHREFFESQA
INTDLSNAMLPISITIPVPEGRLPVGLSFDALGRDALLSRVSAIEQVLGFRVKY
LPHTT"
BASE COUNT 813 a 917 c 1046 g 901 t
ORIGIN 1 bp upstream of EcoRI site.
Query Match 60.0%; Score 15; DB 3; Length 3677;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gcatcgccagcattt 15
|||||
Db 1729 GCATGCCAGCATTT 1715
|||||
RESULT 38
ABO28980/c 5618 bp mRNA PRI 04-AUG-1999
LOCUS Homo sapiens mRNA for KIAA1057 protein, partial cds.
DEFINITION ABO28980
ACCESSION ABO28980.1 GI:5689450
VERSION ABO28980.1
KEYWORDS Homo sapiens brain cDNA to mRNA, clone_lib:pbluescriptII SK plus
clone:hhl1838.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (sites)
AUTHORS Kikuno,R., Nagase,T., Ishikawa,K., Hirose,M., Miyajima,N.,
Tanaka,A., Kotani,H., Nomura,N. and Ohara,O.
TITLE Prediction of the coding sequences of unidentified human genes.
XIV. The complete sequences of 100 new cDNA clones from brain which
code for large proteins in vitro
DNA Res. 6 (3), 197-205 (1999)
JOURNAL 2 (bases 1 to 5618)
MEDLINE 99397452
REFERENCE Ohara,O., Nagase,T. and Kikuno,R.
AUTHORS Direct Submission
TITLE Submitted (17-JUN-1999) to the DDBJ/EMBL/GenBank databases. Osamu
JOURNAL Ohara, Kazusa DNA Research Institute, Laboratory of DNA Technology;
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
(E-mail:cdna@kazusa.or.jp, Tel:+81-438-52-3913,
Fax:+81-438-52-3914)
FEATURES
Source Location/Qualifiers
gene 1..5618
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone_lib="pbluescriptII SK plus"
/tissue_type="brain"
1..2935
/gene="KIAA1057"
<1..2935
/gene="KIAA1057"
/codon_start=2
/product="KIAA1057 protein"
/protein_id="BAAB3009.1"
/db_xref="GI:5689451"
/translation="LADSSPSNLIKIILLKELLSMHHPDPALTKKFDYLPVDSRSSSG"

```


/translation="MSLPTSLPTVFLDRATIPRIHISLPALPFEHHWLEYDACEPQQ
VVERLIAADIVITNREMLIQLPKLIAISATGTNNVDLPACDRDLNIAVNCVQ
GYATRSVPHVAMFALRNLIAGKGOVCEKRGDYSFEOVLAQSDVLSLHPTDITET
GSGALGOATNARALGHVILAEKRGVCEKRGDYSFEOVLAQSDVLSLHPTDITET
RNIISAEIAQNPALLINTGREGVLEOALVDALKRROLAGAGVDVESAEPADMDN
PLIANDLPLNLLTHPVAGWSSSQIQLATILINDISAFMRGEAKRVV"
complement(6222..6959)
/gene="VC2505"
complement(6222..6959)
/note="VC2505"
/note="similar to SP:P39219 GB:U00096 PID:1786244;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="ribosomal large subunit pseudouridine synthase
A"
/protein_id="AAF95647.1"
/db_xref="GI:9657087"
/translation="MAMTEYNPPKDWIDVIYEDHIIAANKPSGLLVGPRVVEHYD
SMWRLVDCPDVVKVVRHLDMSGLIFAKGHMESALKKGFQRLTHKIIYARVWG
VMPQNEGVNLPCCDMPNPKQKVCFEGRDLSLTRYQVKKREKTTVVKRLIYITGRS
HOLRVHMSQLGHPIDVGDLDYAPAEVAFADRLQLHAAELCFYHPRSHWLRTLFVPCDF
YPAEQLIQLHDPEDPKLDYKKLSQD"
complement(7056..9965)
/gene="VC2506"
complement(7056..9965)
/note="VC2506"
/note="similar to GB:U00096 PID:1786245; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="RNA polymerase-associated protein HepA"
/protein_id="AAF95648.1"
/db_xref="GI:9657088"
/translation="MSPALGOWISDTSGLGTGVVVALDARTVITLMFAAERNRYVA
RSDAPTRVIFNVDVDSQSGSLOVEQGVVYTLGTRVDTEESGVALRELFIL
SNQIRFNKQDKLQAGIDRDNFVLYRALTQYQKHSPMGLGCMRAGLIPHOLY
IAHVRGRRHAPRVLDNADEVGLKTIAGMIIHQVLIQGRALRIIVVPTLQHWLVE
MMRNFHFIIDEECEVFEADNPFEQVVICSLDFLRKSRSPFQALAEWDL
LVVDIAHWHPEKPSREYQVIEALAEQTPGVLLTATPEQIGRESHFARLLDAD
RFYDYFAVEEYQAPADAVATLFGSKLSDEAKNKITELLSQDVPLFKALSS
ASEDEIALAQELIDNMDRGTRVLFTRAAIKGFPVVRNVLHLPLETPSYQTTSM
RVAGMLGKLTPEARAKMLYPEIIFEEFEESWQFDSRVNLLKVKAKRSEKI
LVIASASTALQALREREGRATVFEHGMSTIIPDKAAVFAQEGGAOVLCS
IGSEGNFQANQLVMDLPFNDLQRIQGLRDIQCKRDIIDVYVPLTETSQLILA
RWFQEGNFAETQCDGRAVYDAERLIPIAAGGEELEVIIEESAKNKLTKLSQ
EVGRDLLEHNSGGERAQATIAQIKTDGTNLVTFALSDAIGLHQEDGENALV
VTPAHHMVPSPGLPYEGATITPDRTALSRDMHFISWEHPMWQGGIDLLMSEGVG
TCVALLKNAKALPGTILLEVYVVDQAAPKRGISRLFPVSPIRILMDARGNDLSQ
VEFEFNRQLSPVNRHLASKLVSSVQDVHRLITASETAVEPRVSAIREQAQRDMQGS
LNSLERLALUKAVNPINDEIEVLDQIKELTGTYAQAAQQLDLSRLIVVHN"
BASE COUNT 2663 a 2634 c 2514 g 2394 t
ORIGIN

Query Match 60.0%; Score 15; DB 1; Length 10205;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 8 cagcattcgatcct 22
|||||||
Db 6278 CAGCATTTCGATCCT 6264

RESULT 40
AE003860
LOCUS AE003860 10593 bp DNA BCT 06-JUL-2000
DEFINITION Xylella fastidiosa, section 6 of 229 of the complete genome.
ACCESSION AE003860 AE003849
VERSION AE003860.1 GI:9104830
KEYWORDS Xylella fastidiosa.
SOURCE Xylella fastidiosa
ORGANISM Xylella fastidiosa

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS

1 (bases 1 to 10593)
Silvestri,M.L., Siqueira,W.J., de Souza,A.A., de Souza,A.P.,
Terenzi,M.F., Truffi,D., Tsai,S.M., Tshako,M.H., Vallada,H., Van
Sluys,M.A., Verjovsky-Almeida,S., Vettore,A.L., Zago,M.A., Zatz,M.,
Meldanis,J. and Setubal,J.C.
The genome sequence of the plant pathogen Xylella fastidiosa. The
Xylella fastidiosa Consortium of the Organization for Nucleotide
Sequencing and Analysis, Sao Paulo, Brazil
Nature 406 (6792), 151-157 (2000)
2 (bases 1 to 10593)
Simpson,A.J.G., Reinach,F.C., Arruda,P., Abreu,F.A., Acencio,M.,
Alvarenga,R., Alves,L.M.C., Araya,J.E., Bala,G.S., Baptista,C.S.,
Barros,M.H., Bonaccorsi,E.D., Bordin,S., Bove,J.M., Briones,M.R.S.,
Bueno,M.R.P., Camargo,A.A., Camargo,L.E.A., Carraro,D.M.,
Carrer,H., Colauto,N.B., Colombo,C., Costa,F.F., Costa,M.C.R.,
Costa-Neto,C.M., Coutinho,L.L., Cristofani,M., Dias-Neto,E.,
Docena,C., El-Dorri,H., Facincani,A.P., Ferreira,A.J.S.,
Ferreira,V.C.A., Ferro,J.A., Fraga,J.S., Franca,S.C., Franco,M.C.,
Frohme,M., Furlan,L.R., Garnier,M., Goldman,G.H., Goldman,M.H.S.,
Gomes,S.L., Gruber,A., Ho,P.L., Hoheisel,J.D., Junqueira,M.L.,
Kemper,E.L., Kitajima,J.P., Krieger,J.E., Kuramae,E.E., Laigret,F.,
Lambais,M.R., Leite,L.C.P., Lemos,E.G.M., Lemos,M.V.F., Lopes,S.A.,
Lopes,C.R., Machado,J.A., Machado,M.A., Madeira,A.M.B.N.,
Madeira,H.M.F., Matsukuma,A.Y., Menck,C.F.M., Miracca,E.C.,
Miyaki,C.Y., Monteiro-Vitorello,C.B., Moon,D.H., Nagai,M.A.,
Nascimento,A.L.T.O., Netto,L.E.S., Nhani Jr.,A.A., Nobrega,F.G.,
Nunes,L.R., Oliveira,M.A., de Oliveira,M.C., de Oliveira,R.C.,
Palmeri,D.A., Paris,A., Peixoto,B.R., Pereira,G.A.G., Pereira
Jr.,H.A., Pesquero,J.B., Quaggio,R.B., Roberto,P.G., Rodrigues,V.,
de M. Rosa,A.J., de Rosa Jr.,V.E., de Sa,R.G., Santelli,R.V.,
Sawasaki,H.E., da Silva,A.C.R., da Silva,F.R., da Silva,A.M., Silva
Jr.,W.A., da Silva,J.F., Silvestri,M.L.Z., Siqueira,W.J., de
Souza,A.A., de Souza,A.P., Terenzi,M.F., Truffi,D., Tsai,S.M.,
Tshako,M.H., Vallada,H., Van Sluys,M.A., Verjovsky-Almeida,S.,
Vettore,A.L., Zago,M.A., Zatz,M., Meldanis,J. and Setubal,J.C.
Direct Submission
Submitted (02-JUN-2000) Organization for Nucleotide Sequencing and
Analysis, Bioinformatics Lab - IC/Unicamp, C.P. 6176, Campinas, SP
13083-970, Brazil

Location/Qualifiers
1. .10593
/organism="Xylella fastidiosa"
/db_xref="taxon:2371"
complement(277..1356)
/gene="XF0061"
complement(277..1356)
/note="similar to SP:P07674 (percent identity: 28 %/query
alignment coverage: 64.9 %/subject alignment coverage:
65.1 %); identified by sequence similarity; putative; ORF
located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=1
/product="transcriptional repressor"
/protein_id="AAF82874.1"
/db_xref="GI:9104831"
/translation="MLLQTLFELDRATAMSLFCNRIKPTQEQNOLATVAFSGEITEE
IDPELIDLEKMRSKDNPFTIELMELGNDMRQDQHEPVLIRKSKPGRYLMVAG
ERRWLACKAVDIKIKAVIRELTDQVVRQRAENIQRENIQTLEIAVALRDKERLGT
LEKVAENWKGINVSRINFLVETNSLTQASVGETADITANDYLRKVQNT
AAKTIIDQAKVDPGNKRVREKLNQAKPTTNEKNATDSRQESQTFQOEGETHLQ
NISNOEKPAVASAPLPDPAQALVALAAKQMETLCTEVSALSILKHQRQDVISALGDV
LQNVVQAALFTRHVLASLANSK"
complement(1331..1657)
/gene="XF0062"
complement(1331..1657)
/note="hypothetical protein; identified by sequence

Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
Xylella.

```

similarity; putative; ORF located using Glimmer/RBSfinder
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF82875.1"
/db_xref="GI:9104832"
/translation="MYHRSIFORSNSIHNETVHKWSDSQNHDIOTTYPLPNTNNI
LEWQNFKNWEICAENTSPQSIYRPPPEQHVNIFFILPHLMFVQHLKMYQARHA
VVTDIV"
gene
complement(2427..3152)
/gene="XF0063"
CDS
complement(2427..3152)
/gene="XF0063"
/note="similar to GI|5921548 (percent identity: 37 %/query
alignment coverage: 93.4 %/subject alignment coverage:
95.7 %); identified by sequence similarity; putative; ORF
located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="competence protein F"
/protein_id="AAF82876.1"
/db_xref="GI:9104833"
/translation="MKDAVNSNHILPDTWCDRLRLRWLLPSYACNPEGHGKTLTCA
CQLLPNENACLRCAAPKEKEIIEPTGCPCLYDPPLETAHAAPLYHWPNTKLISR
FKFHQDIACLLADNMAAQAHPQPAQVPVSLHHGLRTRGVDALELAHHVGNR
LOLPCLPALRRVRATLPQSQALNNKRNSLNHAFIPHLPAAHVALVDDVMTGTGLH
AAARALRHTGVTRIDAWICARV"
gene
3202..4227
/gene="XF0064"
CDS
3202..4227
/gene="XF0064"
/note="similar to SP|P36569 (percent identity: 60 %/query
alignment coverage: 94.4 %/subject alignment coverage:
93.1 %); identified by sequence similarity; putative; ORF
located using Glimmer/RBSfinder/Start codon shift: 66"
/codon_start=1
/transl_table=11
/product="biotin synthase"
/protein_id="AAF82877.1"
/db_xref="GI:9104834"
/translation="NSSIIRYDWTAEELHAFDLSPLELLYRAASVHRHFDPAEIOV
STLLSVKTCGPCEDSYCPQAOYRGTVTAOKLMDVDVAVAKAQAQKLAGASRCMGA
AWRSPDRDIPKIAAMIREVKALGETCATGLMNTCAQALKDAGLOYIYNNVDTSP
DEVDVSIHTRQYQDRDLRLTAHVDVGLKTCGGIVGMGETRQHRVGLLLYTLTLPHP
DSVPVNLVQVAGTLPHTGTQTDLPFEFVRMTAVARTPMRSMVRLSAGRESMSDELQ
LCFMAGANSIFYGEKLLTANPETERDQALFQRLGLRPMHLMENYNSQDQHGNVHAD
IACKVVV"
gene
4281..4382
/gene="XF0065"
CDS
4281..4382
/gene="XF0065"
/note="hypothetical protein; identified by sequence
similarity; putative; ORF located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF82878.1"
/db_xref="GI:9104835"
/translation="NATLVDVVSFMYDGVNLDSDSGDGERKVVVEFS"
4486..5673
/gene="XF0066"
CDS
4486..5673
/gene="XF0066"
/note="similar to SP|O34731 (percent identity: 33 %/query
alignment coverage: 63.3 %/subject alignment coverage:
96.2 %); identified by sequence similarity; putative; ORF
located using Glimmer/RBSfinder"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF82879.1"
/db_xref="GI:9104836"
/translation="MSLINHKNGVSCMGSTCSRDTILNRLLMWRVVSNGNRYPVHLVS

```

Brody, L.L., Coulter, S.N., Folger, K.R., Kas, A., Larbig, K., Lim, R.,
 Smith, K., Spencer, D., Wong, G.K., Wu, Z., and Paulsen, I.T.
 Complete genome sequence of *Pseudomonas aeruginosa* PA01, an
 opportunistic pathogen
 Nature 406 (6799), 959-964 (2000)
 2043737
 2 (bases 1 to 10978)
 Stover, C.K., Pham, X.-Q.T., Erwin, A.L., Mizoguchi, S.D., Warren, P.,
 Hickey, M.J., Brinkman, F.S.L., Hufnagle, W.O., Kowalik, D.J.,
 Lagrou, M., Garber, R.L., Goltry, L., Tolentino, E.,
 Westbrook-Wadman, S., Yuan, Y., Brody, L.L., Coulter, S.N.,
 Folger, K.R., Kas, A., Larbig, K., Lim, R.M., Smith, K.A., Spencer, D.H.,
 Wong, G.K.-S., Wu, Z., Paulsen, I.T., Reizer, J., Saier, M.H.,
 Hancock, R.E.W., Lory, S., and Olson, M.V.
 Direct Submission
 Submitted (16-MAY-2000) Department of Medicine and Genetics,
 University of Washington Genome Center, University of Washington,
 Box 352145, Seattle, WA 98195, USA
 Location/Qualifiers
 1. 10978
 /organism="Pseudomonas aeruginosa"
 /strain="PA01"
 /db_xref="taxon:287"
 complement(132..449)
 /gene="PA3762"
 complement(132..449)
 /gene="PA3762"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAG07149.1"
 /db_xref="GI:9949932"
 /translation="MYKLCFYVPESHLDVVKQAVFAAGGGRIGAYDSCCWSLQGGQF
 RPLDQSPYLQGVQVEHVAEMKVELVVADELHSHVSKALKAAHPYETPAYEVRWLTDD
 MVF"
 complement(914..4810)
 /gene="purL"
 /note="PA3763"
 complement(914..4810)
 /gene="purL"
 /codon_start=1
 /transl_table=11
 /product="phosphoribosylformylglycinamide synthase"
 /protein_id="AAG07150.1"
 /db_xref="GI:9949933"
 /translation="MLLIURGAPALSAPRHGKLLQLQHVPAVTGLYAEFAHFADVTG
 ALTADQVLAQLKYPSPVQEPGSLFLVPRFGTISPWSKASDIARNCGLAKI
 DRLERGIAYVQGESDQAQVAARLHDMRTQLVLDLEGAELFSAQPRPLTAVD
 VLGGRAALEANVELGLALAEIDYLLKSGELGRNPHDVELMFAQSEHCRHK
 IFNASHDIDQADKSLFKMLKNTYEMREGVLSAYKDAAVIVGHVAGRFPPDQTR
 EYASREPVLQMKVETNHPTAIPFPGASTGSGGEIRDEGATGRGAKPKAGLTGT
 VSNLQIPGEQPEVYQPKERIVTALDIMEVGLPGGAAPFNNRGRPALTYPTFEQ
 KIATPHGEVRYHKPIMLAGMGNIREDHVOKGEISVGAkliVLGGPAMLIILGGGA
 ASSMATGASDLDFASVORNPENRRCQEVIDRCWLGERNPISFTHDVAGGLSN
 ALPELINDGRCGRELRAVPNDPEGSPLEIWCNESQERYVLSYDAADFETFKAIICE
 RERCPPAVGEAIEORLTVDSDHFDNKPVDMPLEVLIGKAPRMHRAVTRAEALGDDF
 DAAGLEQSEVRLRHPAVASKSLITIGDRTITGLVARDQMGPQVOPVADCAVTA
 TSFDTITGAMWANGERTPLALDDPASGMAIGERTYNLAARVGLKSLDKLSANWA
 AAGHPGEDARLYDITKAVGMELCPGLGITIPVGDNSMKTQWQNGEDKSVTSVSL
 IYTGAPVADVRQSLTPQLRDLKGETDILIDLGKKNLGGSLIAQVHGLGRAPVD
 VDDALQFAFFAIVQGLNADGHILAYHDSRDLGLTSVLEMAFAGHCGVELNLDALD
 SREELAAVFEELGAVIQVREGATPEVLAQSAAGLDDCAVIGQPVNGEINLNYD
 GETVYSNORILQVITWSETSYQIQRNRNADCAEQEFDALIDENPGLSIKLSYDND
 DIAPIYIKKVRPKVAILREQVNGQVEMAADFRAFGAIDVHMSDILAGRVDLDAF
 KHLVACGGSYGDVLAGBGGWAKSILFNARNDGQAFQKDGSAFGVNCQMMQSN
 LHLEIPTEGFHFVHRNRESEQFEARVAMVQVQESSIFPQKMGSRPLPIATHGEGHA
 EFSEBALLEDLSCVSLRFDVNDHKGTEAYPAHPNGSPRGITGLSSRDGRVITLMP
 HPRVFRVQNSWRPDDQEDGGWLRMFRNARVVD"
 5211..6569
 /gene="PA3764"
 5211..6569
 /gene="PA3764"
 /codon_start=1

/transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AAG07151.1"
 /db_xref="GI:9949934"
 /translation="MKEGVLRVITRNSPATYFQDRNGETGEFEYLAKRFAERLGLVEL
 KIETADNLDLYAOLSRREGPALAAAGLTTPGREDASVYSHTYLDVTPQIYYRNGQQ
 LITLVDNELANQVYFNVVAFDFGEARGALAWLPGGDDSDLMNEVNAFLDQAKKEG
 LLQURDRIYGHVDVLGVGATFAQHLQOBLPRYPYSHFKOSGKOLDTMDRLAAIGY
 QESLWPGATSKTGVRLMLNRTAQAMGVSRLDPKQSKOSQGGSKYFVQIRSELPEP
 IKEPDRSWFALAAYNIGAHLEDARKMAEKELGNPNKWDVKKMLPLRLAQKQWIAKTR
 YGARGGETHEVQNVRRYYDILTWVTPQMEGSQLAESGLHLPGVNKRTRPEDSGDE
 KI"
 complement(6583..7167)
 /gene="PA3765"
 complement(6583..7167)
 /gene="PA3765"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAG07152.1"
 /db_xref="GI:9949935"
 /translation="MANDPYIQSFCTGDTGTRLLTAALPEVALALKQAVAGCAAGPLL
 VFDNATGRSVDLDIRGSEDLRLVLVAGERQPAEEAPRGGRFLGLVVADEVITLLP
 RHWELAAQPGGASVALRKLVEARRSQSRQAQERAYHFPMNMAAGDLPGFEATRR
 ALFAGDREGFAERIAGWPTDVEHAAWLASDGES"
 complement(7293..8549)
 /gene="PA3766"
 complement(7293..8549)
 /gene="PA3766"
 /codon_start=1
 /transl_table=11
 /product="probable aromatic amino acid transporter"
 /protein_id="AAG07153.1"
 /db_xref="GI:9949936"
 /translation="MSDTRVQVQDHAGAEAVDSSGLEVKRLTFLEAVAMIVGTNIGA
 GVLMSAYSRKAGFMPLLLWAVAGLFTTISMLIVSETALRTHNQLSGLAQRYVGS
 FGWAIFLSVAVNSIGALITVMSGSKILSAFFGISPALGSVLFFIPAAGVLYLGLSA
 IGKGRKFSIGVSMILILVAATLLNDNTEFARLDGDMIVYVFNIAVFCFSAQYI
 VPMARGFSHAPELPAKVTIGMLTTFVLLSIVPLSIALTGLENSEVATLAWKAL
 GEWAFTANTFALCAMLTSYVGLSGSFLTNIFDKFKLGPENRPTKRLVLAVALPP
 FVLAYSGLVGFVNALYFAGAFSGVLLSIMPILMRSARRNGDFEPAWKCGWIAHPAIQ
 AVITYILASAAIAICSAMLLPAGW"
 complement(8938..9486)
 /gene="PA3767"
 complement(8938..9486)
 /gene="PA3767"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AAG07154.1"
 /db_xref="GI:9949937"
 /translation="MRVTKLISNGKGYTVKVRPIIDRSRDPHEMREALAEAKAA
 LGEPVCAVLVREGELTGRGNFPISSHDPSSAHAEMLAIRMAAAEAGNYRLPGSTLYV
 LTFPCSMCSGLLVHARLQRYTVETVEPKCAVESGRFFEQEHLNHRVMVEGGVLAEE
 CSQALSQFFFRAREAAKRPED"
 complement(9477..10868)
 /gene="PA3768"
 complement(9477..10868)
 /gene="PA3768"
 /codon_start=1
 /transl_table=11
 /product="probable metallo-oxidoreductase"
 /protein_id="AAG07155.1"
 /db_xref="GI:9949938"
 /translation="MTFTROVLGLAGLAVVGLGAGGARLWLARPOVAQVQYDELTA
 APLDLIYVPGSSPALAGGCPGCVELRAKQWELRVRTNRLDEPTTIHWHGIRLPI
 EDMGVPTISQPPVQGESFIYQFQDAGSVYHPHLMSSQLGVLGLPIIEREP
 TCFRHEKVLCLTKTHVDEQATFEPFQAARREGTGRYSTINCKHVPTDLPAGOI
 VVRLLNVDNTVTYRLNLPNGEARIYIDGHPVEPRFEGQIWIQPNRLELALKVPE
 ATELSDRGVPRLATIRSAEAPAGDMWPKPLPANPVSSEPLDANAEKIGFPEWVG
 AMSDTSGKNPYSPFWQINGKAWEGEEHKNAPLAKLKEQSYIFELRNNAQYOHPI
 HLHMGAFVLDSDRDIIPYFTDTYLLCKNETARVALVDNPLGLMFMFCHVIDHMETG

```

BASE COUNT      1876 a 3658 c 3644 g 1800 t
ORIGIN
    LMGTIATGCAWCG"
Query Match      60.0%; Score 15; DB 1; Length 10978;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 atcgccagcattcg 17
|||||
Db 9239 ATCGCCAGCATTTCG 9253

RESULT 42
AC018333/c
LOCUS
DEFINITION      AC018333 23873 bp DNA HTG 09-DEC-1999
                  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
                  pieces.
ACCESSION      AC018333
VERSION        AC018333.1 GI:6552858
KEYWORDS       HTG; HTGS_PHASE2.
SOURCE         fruit fly.
ORGANISM       Drosophila melanogaster
                  Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
                  Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
                  Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE      1 (bases 1 to 23873)
AUTHORS        Adams,M. and Venter,J.C.
TITLE          Direct Submission
COMMENT        Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
                  Rockville, MD, USA
                  This sequence was identified as CDM:10214399 by the submitter.
                  * NOTE: This is a 'working draft' sequence.
                  * This sequence will be replaced
                  * by the finished sequence as soon as it is available and
                  * the accession number will be preserved.
FEATURES       Location/Qualifiers
                  source
                    1..23873
                    /organism="Drosophila melanogaster"
                    /db_xref="taxon:7227"
BASE COUNT      6912 a 4665 c 4568 g 7728 t
ORIGIN
    1 gcatcgccagcattt 15
    |||||||
    Db 2224 GCATCGCCAGCATT 2210

Query Match      60.0%; Score 15; DB 65; Length 23873;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcatcgccagcattt 15
|||||
Db 2224 GCATCGCCAGCATT 2210

RESULT 43
AF168792/c
LOCUS
DEFINITION      AF168792 31424 bp DNA VRL 23-SEP-1999
                  Gallid herpesvirus 1 ICp18.5 gene, partial cds; DNA binding
                  protein, DNA polymerase, UL31 homolog, UL32 homolog, UL33 homolog,
                  UL34 homolog, UL35 homolog, very large tegument protein, UL37
                  homolog, capsid assembly protein, ribonucleotide reductase large
                  subunit, ribonucleotide reductase small subunit, virion host
                  shut-off (VHS), and UL42 homolog genes, complete cds; and UL43
                  homolog gene, partial cds.
ACCESSION      AF168792
VERSION        AF168792.1 GI:5918968
KEYWORDS       gallid herpesvirus 1.
SOURCE         gallid herpesvirus 1.
ORGANISM       Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
                  Alphaherpesvirinae; Varicellovirus.
REFERENCE      1 (bases,1 to 31424)

```

AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source

Johnson, M.A.
Sequence of the infectious laryngotracheitis virus (SA-2 strain)
unique long region UL28 to UL43
Unpublished
2 (bases 1 to 31424)
Johnson, M.A.
Direct Submission
Submitted (14-JUL-1999) Division of Animal Health, CSIRO, Private
Bag 24, Geelong, VIC 3220, Australia
Location/Qualifiers
1..31424
/organism="gallid herpesvirus 1"
/strain="SA-2"
/db_xref="taxon:10386"
complement(<1..752)
/note="UL28 homolog"
/codon_start=1
/product="ICp18.5"
/protein_id="AAD56200.1"
/db_xref="GI:5918969"
/translation="MAETQLRLMALAGLOLTLLFQIDILKRCDEPILLOKRLKRSRIK
HNALMALYIHSLLSNEISMQVNYRLTQTYCWLMLRRACSEIAALCNEIDKYSVLKQ
RKCFEATLGYGKGFICPLHAEVQLQIIGMSIIISKELVINDVENFMKQINCHV
SCKNAIEALNTIDEFLVSSGSGSLVAAPDYDHGQPCVVCLEELSVTANOGETIYRRL
GYKICDHLTRQFPVNVSTDDILKRLHLPFLRDVDP"
complement(800..805)
complement(834..3833)
/note="UL29 homolog"
/codon_start=1
/product="DNA binding protein"
/protein_id="AAD56201.1"
/db_xref="GI:5918970"
/translation="MEKSSGCTTGGKSDGCPWCIIGVYVYAPREIDGKEWKLIC
AKSDQSCATAPLIRGLTVESDFKPNVAAVIGTKSGVGVGGNCTAILSPCHFTVY
IFHGECIPPTFTPLNTKICEDAREFGFSSPPNGPVPNAIETGEETCKSLNMDP
DKMLYLVAEPFCEAVCYNTIYHGSVETLYINSHVTRIPITYPQVMPDIALRI
CRNPFDSNRNIGECIYPTPLFNKALNRLIHGATLGTQGSRLTROLEAVARGAAML
AFDGFEGCVLTSDKTELAANOQTQTPPKQNVVEERRAACSLAALALATRLSVSC
APYDFDNCGLPYEWPFLCDCKTOMERVALDFSAELAGIAGVGAALFSSNNPLIASE
VVDGAPPGERASSGLTRFLYTCGLHLLGCPQVDYSGKRWGCTPSPUTPONTPEY
DPEHLAYACGFPSELLRIIFYLEKSKQLSTRPNMNLISISNSGKAPCTGAVPEY
CNDESRNICRYTMQRTRFPPIPTMRRSPIAVIGAVDGDYDCTDILGNFAYSQL
RAGDEPAKAVMNDTYRIGWRVQYLVSEGLNKDGTEDHNTISLAEKLTVEKIQ
NFVGESECKFISALSGVRYHYKHELSIIVHTFGSPPEYSSAYCPVLSLCAOTRSL
FQDLILSQIHGTFTDTPROESKMPFRRSAMPALRAFNGMLDKGLSGKYEPFSISVPCV
NAPDTMRPNTQALAOEYSISRGVYLKKEFKVKNRVFNGSSTNSCRVLSRRPI
VLMGIMISKVVGQONSTTFQSGNWGSLICNSGQSVNSALTDNPKKFAACAKRGI
ILSGGVASQFQENTIASQVRSLLDAGGSPTSIYAVLRVLGEGMDITAEITWMAITD
DKLVNLVLBELDISGQNGSVATAQSMLELEKDKTGPENAGEMLVWADEDTYN
TGDISEDIELGAPSAKKPSLSADILFC"
4010..7033
/note="UL30 homolog"
/codon_start=1
/product="DNA polymerase"
/protein_id="AAD56202.1"
/db_xref="GI:5918971"
/translation="MDTFFSAGTGAARVAPIILGLPHSGASYTYSVREFFYVCPTCIN
GGRIQTQGRATSKPTFTHNERQDLITSTHGAWPVRMYKNAGPMTHHNKGDYKF
YEFVYDLDETEIAQSCVRMHSEFLDTLKPTGTVITLVGNSACKRVAVHVGSLP
YFVKRREIDQAAKVTNSBELAHALATKTSKNSPFLVSAESFCDIVVRRRDY
YSESEEDYRVKVCNGKVMKTCIDNFPSPKYESNVDAITREILDNLTSGWYRF
RAOGGALOIRDPQGHATSADVELINCTASNLGNVSWPDYKLLCFEDICKAGANEF
AFPSAEKVELVQISAVYSLTKEKEGIIIFSGTCSELSDCSDGTVTCAGSFE
LLICPMFFQYSPERVSGYNILGDMAYLSNKLKSKVGMRLDYGKSGEAKWTFKLD
PSARGLGRFKKKNVINGVNDMPTISYKELKLPVSKLVNAVECVGEQKIDLAYKIDP
VMPAAGPKERKIGEYCLQDSRLSGSLFFKPHLEMSAVAKLACIPLTRALDGGOL
RVYTCLLQRSTAGFVLPDKKSAFSGSTLASDAPATRNVGAKVLDPEIGFH
VTPVVYDFEASLYPSIIQSNNLCTLTHDPAALAGLOPEKDFDIDVQGRFVRE
HIRKSLDILGLDLSMRKAVRAKIKTAETEEERILLDKQAALSCLEICGYFTGMH
GMPLCEVASTVTAIGRMLLTKAHIEKWRSGNQAFLPFGSERIQOLNYSVRVI
NGETDSAYCEYVRSRLHNFNSWHLGRINDLNKRLFLVKRLKLEKFSACLLIGKEK
YIGVHHGKMLMKRVLDVLFKTKNFVNTASRLVNLFLDEDEIAMAADTARAGFDFD
CLPSGLTKLGLRTAEARLAIITGNLNIIRDIFIMTAELSRAVDNYASLKLPHTVYQKKA

polya_signal
CDS

CDS

JOURNAL	Submitted (12-SEP-2000) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
REFERENCE	5 (bases 1 to 94487)
AUTHORS	Town,C.D. and Kaul,S.
TITLE	Direct Submission
JOURNAL	Submitted (05-OCT-2000) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
REFERENCE	6 (bases 1 to 94487)
AUTHORS	Town,C.D. and Kaul,S.
TITLE	Direct Submission
JOURNAL	Submitted (12-OCT-2000) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
COMMENT	On Dec 10, 1999 this sequence version replaced gi:6143858. Address all correspondence to:at@tigr.org
BAC clone F15M4 is from Arabidopsis thaliana chromosome 1 The orientation of the sequence is from SP6 to T7 end of the BAC clone. Genes were identified by a combination of several methods: Gene prediction programs including Genscan+ (Chris Burge, http://CCR-081.mit.edu/GENSCAN.html), GeneMarkHM (Mark Borodovsky, http://genemark.biology.gatech.edu/GeneMark/), GlimmerA (a variant of GlimmerM, see Mihaela Pertea, http://www.tigr.org/softlab/glimmerm.htm/glimmerm.html, and GeneSplicer (Mihaela Pertea and Steven Salzberg, contact mpertea@tigr.org), searches of the complete sequence against a peptide database and the plant EST database at TIGR (http://www.tigr.org/tdb/tgi.shtml). Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as unknown proteins. Genes without protein or EST similarity, that are predicted by more than two gene prediction programs over most of their length are annotated as hypothetical proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats are identified by RepeatMasker (Arian Smit, http://ftp.genome.washington.edu/RM/RepeatMasker.html).	
FEATURES	Location/Qualifiers
source	1..94487 /organism="Arabidopsis thaliana" /cultivar="Columbia" /db_xref="taxon:3702" /chromosome="1" /clone="F15M4" complement(967..1017) /rpt_family="AT_rich" complement(1289..1338) /rpt_family="AT_rich" complement(4053..4080) /rpt_family="AT_rich" complement(join(<6943..7515,7607..7953,8115..8296,8522..>9510)) /gene="F15M4.1" complement(6943..9510) /gene="F15M4.1" /note="identical to hydroxy methylglutaryl CoA reductase (AA 1-592) GB:X15032 [Arabidopsis thaliana]" complement(join(7255..7515,7607..7953,8115..8296,8522..9510)) /gene="F15M4.1" /codon_start=1 /product="hydroxy methylglutaryl CoA reductase (AA 1-592); 9510-7255" /protein_id="AAF16652.1" /db_xref="GI:6554470" /translation="MDLRRPDKPPVTNNNSGSRFSYOPRTSDDHRRATTIAPP PKASDALPLYLNTAVFFTLFFSVAYLLHRRWPKIRYNPLHVVTITELGATIALI ASFTYLLGFFGIDFVQSFISRASGDAMDLDITDDHRLVTCSPPTPIVSVAKLPNP EPIVTESLPEDEEIVKSVIDGVPISVSLSGLDCKRAASIRREALQVTVGRSIEGL PLDGFDESILGQCCMPGVYIQIPVGITAGLLDGYEYVPMATTEGLVAVFNRSR KAMPTSGGATSTVLKDGTRAPVVRFAASRASELKTFLENPENFDTLAVVFNRSR ARLOSVKCTIAGKNAYVRFCCSTGDGMNMVSKGVQNVLEYLTDDFPDMVDVIGISGN
repeat_region	
repeat_region	
repeat_region	
mRNA	
gene	
CDS	

```

FCSDKPAANVNIIEGRGKSVVCEAVIRGEIVNKVLTSSAALVELNMLKLNAGSAVAG
SLGFGNASHASIVSAFVIAIGQDPAONVSSOCITMNEAINDGKDHIHSTYMPISIEVG
TVGGTOLASQACNLILGVKGASTESPCHNARRLATIAGAVLAGELSLSAIAAQQ
LVRSHMYNRSRRDISATTTTTTTT
complement(7258..7283)
/rpt_family="(CAA)n"
complement(11154..11209)
/rpt_family="AT-rich"
complement(11551..11580)
/rpt_family="AT-rich"
complement(11933..12004)
/rpt_family="AT-rich"
join(<12390..12391,12491..12629,12730..12861,12929..13019,
13099..>13268)
/gene="F15M4.2"
12390..13268
/gene="F15M4.2"
/note="similar to unknown protein GB:AAD30603 [Arabidopsis
thaliana]"
join(12390..12391,12491..12629,12730..12861,12929..13019,
13099..13268)
/gene="F15M4.2"
/codon_start=1
/product="unknown protein; 12390-13268"
/protein_id="AAF16653.1"
/db_xref="GI:6554471"
/translation="MEDVKALXVKGELRETLPQSGSRVYKQLGKNSWYEVKISY
PASIPARVSLQILKHEMELKLNMRLLNTEKLIIFKAESLKEVNKAGLNVLPLEP
EGIVAPNSRERYFIYINIVCEQLMGIPYSSVSVLLVALCLVSVFVPSRLPSLL
TKDQGLRSHRSHGKDS"
join(<14056..14170,14277..14431,14516..14701,14819..14990,
15096..15285,15362..>15506)
/gene="F15M4.3"
14056..15506
/gene="F15M4.3"
/note="similar to cinnamoyl-CoA reductase GB:CAA56103
[Eucalyptus gunnii]"
join(14056..14170,14277..14431,14516..14701,14819..14990,
15096..15285,15362..15506)
/gene="F15M4.3"
/codon_start=1
/product="putative cinnamoyl-CoA reductase; 14056-15506"
/protein_id="AAF16654.1"
/db_xref="GI:6554472"
/translation="MAYQKVCVTCAGGFIAISWLVKFLLSRGYTVHGTVRDPCDEKND
HURKLNASKNLKFKADLFDDEGLFSAIDCGSVFHIAFPVEGELIKPALTGK
NVLEATEKTVQVSVSIAAVYNPKVPQDVARDECDSDTQVTLHSELEGWRYVL
AKTLTERALEWSKRNFAVVTLCPSVIGPRLSTLNSSSLGLKFKIGKIKSLSD
ELYLDVRDVAALLLVYENREATGRYICNSHSLYTDLSMEKLNMYPRKRPESFTE
VKEKEVRPLSAEKLKNGKFRPLEETIDDSVSVSFEAAGDLFKA"
16281..16361
/rpt_family="(CAT)n"
join(<16955..17161,17385..17529,17639..17755,17844..17884,
17978..18215,18332..>18423)
/gene="F15M4.4"
16955..18423
/gene="F15M4.4"
/note="contains Pfam profile: PF00076 RNA recognition
motif. (a.k.a. RRM, RBD, or RNP domain)"
join(16955..17161,17385..17529,17639..17755,17844..17884,
17978..18215,18332..18423)
/gene="F15M4.4"
/codon_start=1
/product="putative RNA-binding protein; 16955-18423"
/protein_id="AAF16655.1"
/db_xref="GI:6554473"
/translation="MAYQVPVSGSFHYLNSPFGDTTFTKVFVGLAWETQSETLRQHF
EQYGEILEAVYIADKNTGRSGYGFVTFROPEARRACADPTPIIDRRANCLASIG
RPRPLPVAVIPNPNPDLPTLEMCVLEHLEIIRISNHFHTTSKELSIIMVTAY
GPEYMSQSGLSYPYMGQQLQVYVPGAVNSPVYQGLSQITIPNGHYTAVQGYIS
VPGSHILQLGGPTVSTMTSSMPALQAPYPSGIPGAPVQSHIIVHSPQMQSTASDQ
TTR"
complement(join(<19352..19378,19455..19499,19581..19699,

```

```

19777..19852,19942..20024,20291..20356,20433..20536,
20615..20690,20777..>20843))
/gene="F15M4.5"
complement(19352..20843)
/gene="F15M4.5"
complement(join(19352..19378,19455..19499,19581..19699,
19777..19852,19942..20024,20291..20356,20433..20536,
20615..20690,20777..20843))
/gene="F15M4.5"
/codon_start=1
/product="unknown protein; 20843-19352"
/protein_id="AAF16656.1"
/db_xref="GI:6554474"
/translation="MAAISPWLSSPQSFNSPRVITITRRRCSSISAAISVLDSSNEEQ
HRISSRDHVGMKRRDVMLOIASVVFFLPLALSPAFETNASEAFRVYTDENKFEISI
PQVSIATITGLGPDFTRMESFGKVEAFETLVSLGDRSMQKPVGYTAKLIDSRASKGY
YIEVTLQNPGEARKHLYSAIGMATNGWYNRLYTVTGQFTDEESAQSSKIQKTVKSR
FI"
complement(20998..21026)
/rpt_family="AT-rich"
complement(21140..21167)
/rpt_family="AT-rich"
join(<22462..22639,22837..>23090)
/gene="F15M4.6"

repeat_region
repeat_region
mRNA

Query Match          60.0%; Score 15; DB 12; Length 94487;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

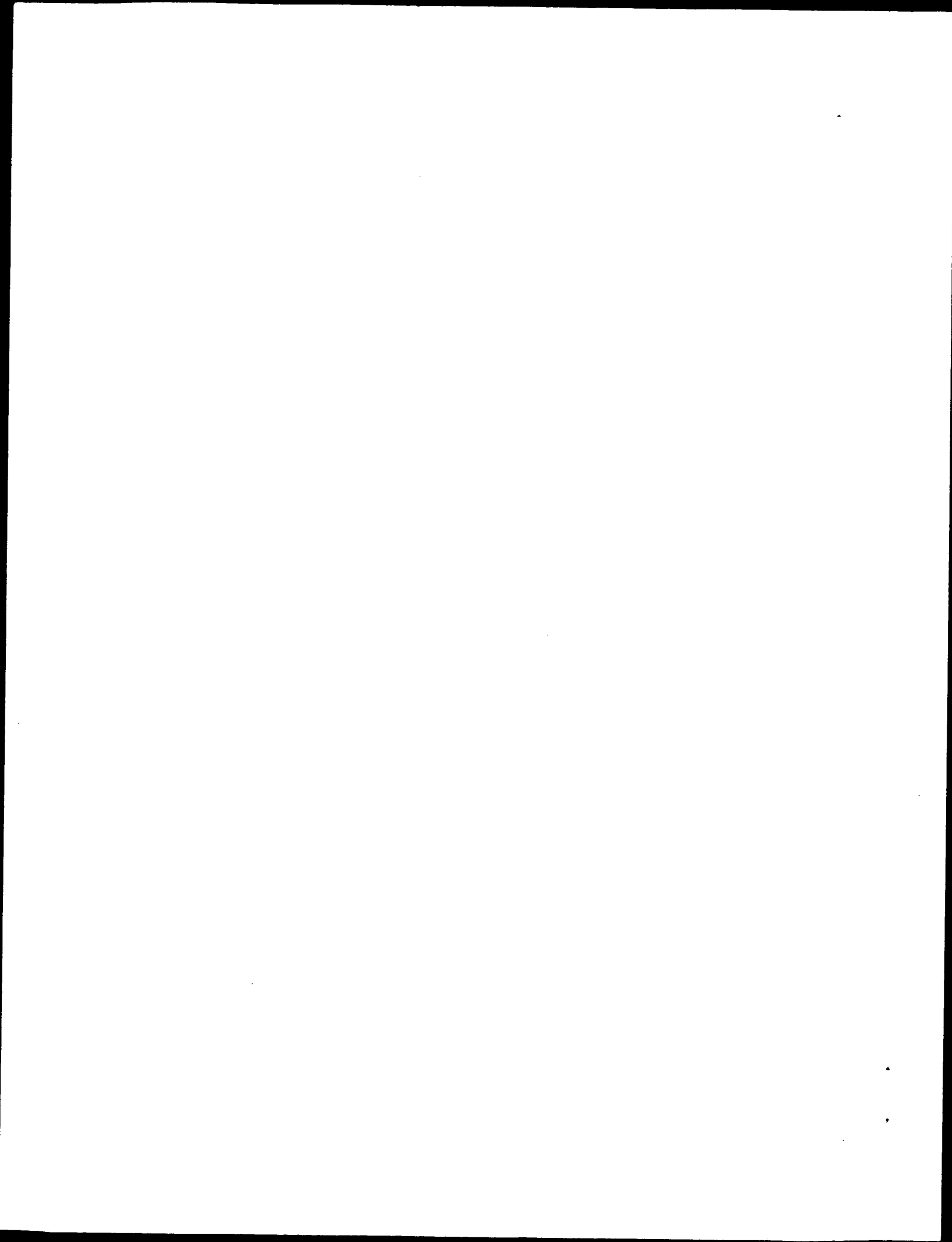
QY 4 tcgccagcatttcga 18
      |||||
Db 17074 TCGCCAGCATITCGA 17088

Search completed: October 9, 2001, 15:47:19
Job time: 14825 sec

```


us-09-396-196f-7.oli.rge

Wed Oct 10 07:46:02 2001



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 15:52:31 ; Search time 469.84 Seconds
(without alignments)
33.410 Million cell updates/sec

Title: US-09-396-196f-7
Perfect score: 25
Sequence: 1 gcatcgccagcatttcgatctcgt 25

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 730101 seqs, 313950809 residues

Word size : 9

Total number of hits satisfying chosen parameters: 24498

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_0601.*
1: /SIDSI/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	100.0	1041	20 AAX01303	E. coli biotin syn
2	25	100.0	1084	10 AAN91329	E.coli Bio B gene.
3	25	100.0	1121	7 AAN60496	Sequence encoding
4	25	100.0	5872	15 AAO62386	Biotin-biosynthesi
5	15	60.0	5773	21 AAT76014	Human OREF ORF1569
6	14	56.0	487	21 AAC74884	Human OREF ORF439
c 7	14	56.0	549	21 AAC93854	Cat flea hindgut a
c 8	14	56.0	1344	21 AAZ56381	Escherichia coli f
c 9	14	56.0	4189	21 AAZ49334	Murine multidrug r
10	14	56.0	4313	14 AAO38950	Mouse multidrug re
c 11	14	56.0	5319	19 AAV35235	R. prowazekii S-la

c 12	14	56.0	143068	21 AAF21105	Human low adenosin
c 13	14	56.0	143068	21 AAF21272	Human low adenosin
c 14	14	56.0	143068	21 AAA34983	Human adenosine re
c 15	14	56.0	143068	21 AAA35150	Human adenosine re
c 16	14	56.0	149412	21 AAA35151	Human adenosine re
c 17	14	56.0	152740	21 AAF21273	Human low adenosin
c 18	13	52.0	22	19 AAV11150	Oligonucleotide #3
c 19	13	52.0	51	21 AAA76686	Human clone cg2838
c 20	13	52.0	51	21 AAA76687	Human clone cg2838
c 21	13	52.0	61	20 AAZ19606	Complement system
c 22	13	52.0	300	20 AAZ14744	Human gene express
c 23	13	52.0	322	21 AAA31174	Plant microsatelli
c 24	13	52.0	447	17 AAT27962	Hepatitis C virus
c 25	13	52.0	591	21 AAC54077	Arabidopsis thalia
c 26	13	52.0	606	21 AAF09259	Arabidopsis thalia
c 27	13	52.0	629	21 AAF09259	Arabidopsis thalia
c 28	13	52.0	714	20 AAZ17005	Fusarium venenatum
c 29	13	52.0	725	22 AAF22436	Human gene express
c 30	13	52.0	750	20 AAZ17191	Human breast cance
c 31	13	52.0	974	21 AAC35170	Human gene express
c 32	13	52.0	974	21 AAC48947	Arabidopsis thalia
c 33	13	52.0	985	21 AAC46920	Arabidopsis thalia
c 34	13	52.0	996	21 AAC47663	Arabidopsis thalia
c 35	13	52.0	1007	19 AAZ96443	Arabidopsis thalia
c 36	13	52.0	1038	21 AAF12052	S. pneumoniae deri
c 37	13	52.0	1065	20 AAX84999	Aspergillus oryzae
c 38	13	52.0	1110	18 AAT84197	Human secreted pro
c 39	13	52.0	1114	21 AAC41953	DNA encoding a UDP
c 40	13	52.0	1179	21 AAC41961	Arabidopsis thalia
c 41	13	52.0	1226	21 AAC36239	Human protein clon
c 42	13	52.0	1263	9 AAN82205	Arabidopsis thalia
c 43	13	52.0	1287	21 AAN93673	Nad B gene encodin
c 44	13	52.0	1291	21 AAN93674	Human peroxisome a
c 45	13	52.0	1313	21 AAC61879	Human peroxisome a
					CDNA encoding a hu

ALIGNMENTS

RESULT 1
AAX01303
ID AAX01303 standard; DNA; 1041 BP.
XX
AC AAX01303;
XX
XX 12-APR-1999 (first entry)
DT
DE E. coli biotin synthetase (BioB) coding sequence.
XX
XX DAP aminotransferase; diaminopelargonic acid; transgenic plant;
KW biotin synthase; biotin production; vitamin H; BioB; ss.
KW
OS Escherichia coli.
XX
XX US5869719-A.
XX
XX 09-FEB-1999.
PD
XX 30-APR-1997; 97US-0846338.
PF
XX 30-APR-1997; 97US-0846338.
PR
XX 08-MAR-1995; 95US-0401068.
PR
XX (NOVS) NOVARTIS FINANCE CORP.
FA
XX Patton DA;
PI
XX WPI; 1999-152902/13.
DR
XX P-PSDB; AAW73906.
DR
XX Transgenic plants with high biotin levels - transformed with DNA
PT encoding di:amino-pelargonic acid amino-transferase or biotin
PT synthase

```

XX PS Example 2; Column 37-40; 34pp; English.
XX CC This sequence encodes the E. coli biotin synthetase (BioB). The gene can
XX CC be used in the transgenic plant of the invention. The transgenic plant,
XX CC plant cell or plant tissue is transformed with a chimeric gene encoding
XX CC diaminopelargonic acid (DAP) aminotransferase or biotin synthase and
XX CC produces more biotin than a non-transgenic plant, cell or tissue. The
XX CC plant is used as an improved dietary source of biotin (vitamin H) for
XX CC humans or animals.
XX SQ Sequence 1041 BP; 262 A; 273 C; 305 G; 201 T; 0 other;

Query Match      100.0%; Score 25; DB 20; Length 1041;
Best Local Similarity 100.0%; Pred. No. 2.2e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcatcgccagcattcgatctcgt 25
   |||||
Db 90 gcatcgccagcattcgatctcgt 114

RESULT 2
AAN91329
ID AAN91329 standard; DNA; 1084 BP.
XX AC
XX AAN91329;
XX DT 15-FEB-1990 (first entry)
XX DE E.coli Bio B gene.
XX KW E.coli; Bio B gene; biotin.
XX OS Escherichia coli.
XX FH Key
XX FT Location/Qualifiers
XX FT CDS
XX FT 24..1064
XX FT /*tag= a
XX PN GB2216530-A.
XX PD 11-OCT-1989.
XX PF 17-MAR-1989; 89GB-0006210.
XX PR 22-MAR-1988; 88GB-0006804.
XX PR 17-MAR-1989; 89GB-0006210.
XX PA (UKAG-) UK MIN. AGRIC. FISH.
XX PI Pearson BM, McKee RA;
XX WPI: 1989-295085/41. P-PSDB P91392
XX Plasmid contg. gene(s) for expression of biotin synthetase enzymes
XX PT - derived from E.coli and capable of replication and expression in other
XX PT microorganisms, esp. yeast.
XX PS Table 3; page 33-4; 52pp; English.
XX CC The gene can be used in a plasmid for expression of enzymes of the biotin
XX CC synthetic pathway. Pref. control sequences for expression in S.cerevisiae
XX CC are plasmids pMA91, pMA36c, pKV49 and pCK495, and plasmid pCK965 for
XX CC Lactobacillus. Insertion of bio B improves biotin yields in
XX CC microorganisms which export biotin, or enables growth in media contg.
XX CC little or no biotin of organisms unable to synthesise biotin for their
XX CC own use.
XX SQ Sequence 1084 BP; 271 A; 286 C; 318 G; 209 T; 0 other;

Query Match      100.0%; Score 25; DB 10; Length 1084;
Best Local Similarity 100.0%; Pred. No. 2.2e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcatcgccagcattcgatctcgt 25
   |||||
Db 113 gcatcgccagcattcgatctcgt 137

RESULT 3
AAN60496
ID AAN60496 standard; DNA; 1121 BP.
XX AC
XX AAN60496;
XX DT 17-OCT-1991 (first entry)
XX DE Sequence encoding biotin synthesising enzyme.
XX KW Biotin synthetic enzyme; E.coli; desthiobiotin; ds.
XX FH Key
XX FT Location/Qualifiers
XX FT CDS
XX FT 42..1082
XX FT /*tag= a
XX PN JP61149091-A.
XX PD 07-JUL-1986.
XX PF 24-DEC-1984; 84JP-0272605.
XX PR 24-DEC-1984; 84JP-0272605.
XX PA (NIPS) NIPPON SODA KK.
XX DR WPI: 1986-216622/33.
XX DR P-PSDB; AAP60536.
XX PT Double stranded DNA encoding biotin synthesising enzyme -
XX PT comprises transformed mutant E.coli strain contg. cyclic doubled
XX PT stranded DNA encoding biotin synthetic biotin enzyme as a plasmid.
XX PS Disclosure; Page 534; 23pp; Japanese.
XX CC The sequence may be expressed by a transformed E.coli host, cultured
XX CC in a medium containing desthiobiotin.
XX SQ Sequence 1121 BP; 290 A; 301 C; 319 G; 211 T; 0 other;

Query Match      100.0%; Score 25; DB 7; Length 1121;
Best Local Similarity 100.0%; Pred. No. 2.2e-05;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcatcgccagcattcgatctcgt 25
   |||||
Db 131 gcatcgccagcattcgatctcgt 155

RESULT 4
AAQ62386
ID AAQ62386 standard; DNA; 5872 BP.
XX AC
XX AAQ62386;
XX DT 16-NOV-1994 (first entry)
XX DE Biotin-biosynthesis genes contg. plasmid pB030A-15/9.
XX KW Biotin; expression; enterobacteria; vitamin H; synthesis;
XX KW plasmid; pB030A-15/9; bioB; bioF; bioC; bioD; bioA;
XX KW promoter ptac; biotin synthase; KAPA synthase;
XX KW 8-amino-7-oxononanoate synthase; pimeloyl-CoA; DTB synthase;

```

KW dethiobiotin synthase; DAPA synthase;
 KW S-adenosyl-L-methionine: 8-amino-7-oxononanoate aminotransferase;
 XX seborrhea; dermatitis; ds.
 OS Escherichia coli DSM498.
 XX

Key Location/Qualifiers
 FH promoter 1..96
 FT /*tag= a "promoter ptac"
 FT /evidence= EXPERIMENTAL
 FT 23..28
 FT /*tag= b
 FT /standard_name= "promoter ptac"
 FT 45..50
 FT /*tag= c
 FT /evidence= EXPERIMENTAL
 FT /standard_name= "promoter ptac"
 FT 105..109
 FT /*tag= d
 FT /evidence= EXPERIMENTAL
 FT /standard_name= "bioB RBS no. 9"
 FT 117..1157
 FT /*tag= e "biotin synthase"
 FT /product= "biotin synthase"
 FT /evidence= EXPERIMENTAL
 FT /gene= "bioB"
 FT /number= 1
 FT 1141..1146
 FT /*tag= f
 FT /standard_name= "bioF RBS"
 FT 1154..2311
 FT /*tag= g
 FT /EC_number= 2.3.1.47
 FT /product= "KAPA synthase"
 FT /evidence= EXPERIMENTAL
 FT /gene= "bioF"
 FT /number= 2
 FT /standard_name= "8-amino-7-oxononanoate synthase"
 FT 2284..2288
 FT /*tag= h
 FT /standard_name= "bioC RBS"
 FT 2295..3030
 FT /*tag= i
 FT /function= "involved in pimeloyl-CoA synthesis"
 FT /product= "protein"
 FT /gene= "bioC"
 FT /number= 3
 FT 3030..3033
 FT /*tag= j
 FT /standard_name= "bioD RBS"
 FT 3043..3753
 FT /*tag= k
 FT /EC_number= 6.3.3.3
 FT /product= "DPA synthase"
 FT /evidence= EXPERIMENTAL
 FT /gene= "bioD15"
 FT /number= 4
 FT /standard_name= "dethiobiotin synthase"
 FT 3712..3750
 FT /*tag= l
 FT /note= "bioD15 substitution"
 FT 3742..3746
 FT /*tag= m
 FT /standard_name= "bioA RBS"
 FT 3750..5039
 FT /*tag= n
 FT /EC_number= 2.6.1.62
 FT /product= "DAPA synthase"
 FT /evidence= EXPERIMENTAL
 FT /gene= "bioA"
 FT /number= 5
 FT /standard_name= "S-adenosyl-L-methionine: 8-amino-

FT RBS 5088..5093
 FT /*tag= o
 FT /standard_name= "ORFI RBS"
 FT 5098..5574
 FT /*tag= p
 FT /function= "unknown, involved in biotin synthesis"
 FT /product= "protein"
 FT /evidence= EXPERIMENTAL
 FT /gene= "ORFI"
 FT /number= 6
 FT 5583..5644
 FT /*tag= q
 FT /standard_name= "rho-independent transcriptional
 FT terminator"
 FT 5583..5605
 FT /*tag= r
 FT stem_loop
 FT W09408023-A.
 XX PN
 XX 14-APR-1994.
 XX PD
 XX 01-OCT-1993; 93WO-EP02688.
 XX PF
 XX 02-OCT-1992; 92CH-0003124.
 XX PR
 XX 15-JUL-1993; 93CH-0002134.
 XX PA (LONZ) LONZA AG.
 XX BI
 XX Birch O, Brass J, Fuhrmann M, Shaw N;
 XX WPI: 1994-135587/16.
 XX DR P-PSDB; AAR51883, AAR51884, AAR51885, AAR51886, AAR51887, AAR63121.
 XX DR
 XX Biotechnological biotin prodn. using enterobacterial biotin-gene
 PT - providing vitamin H in high yield
 PT
 XX Claim 1; Fig 6, Page 47-55 and 60-65; 92pp; German.
 XX PS
 XX The sequence is derived from plasmid pB030A-15/9 contg. the
 CC bioB, bioF, bioC, bioD and bioA genes responsible for biosynthesis
 CC of biotin, arranged in a transcription unit. Microorganisms
 CC contg. these DNA fragments or plasmids may be used in the prodn.
 CC of biotin. Biotin (Vitamin H) may prevent seborrhea, dermatitis,
 CC loss of appetite and tiredness.
 XX
 XX Sequence 5872 BP; 1318 A; 1552 C; 1695 G; 1307 T; 0 other;

Query Match 100.0%; Score 25; DB 15; Length 5872;
 Best Local Similarity 100.0%; Pred. NO. 2.1e-05;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcacgcgcagcattcgatctcgt 25
 |||||
 Db 206 gcacgcgcagcattcgatctcgt 230

RESULT 5
 AAC76014/C
 ID AAC76014 standard; cDNA; 5773 BP.

XX AAC76014;

XX 08-FEB-2001 (first entry)

XX Human ORFX ORFI569 polynucleotide sequence SEQ ID NO:3137.

XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
 KW vulnery; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antinflammatory;

KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
 KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KW thrombosis; contraceptive; ss.
 XX Homo sapiens.
 OS
 PN WO200058473-A2.
 XX
 PD 05-OCT-2000.
 XX
 PF 31-MAR-2000; 2000WO-US08621.
 XX
 PR 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 XX
 PA (CURA-) CURAGEN CORP.
 PI Shimkets RA, Leach M;
 XX
 DR WPI: 2000-602362/57.
 DR P-PSDB; AAB41805.
 XX
 PT Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 XX
 PS Claim 5; Page 2354-2357; 5507pp; English.
 XX
 CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vulnary;
 CC antipsoriatic; antiparkinsonian; neurotropic; neuroprotective;
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotrophic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
 CC antithyroid; and antianaemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.
 XX
 SQ Sequence 5773 BP; 1691 A; 1125 C; 1269 G; 1687 T; 1 other;
 Query Match 60.0%; Score 15; DB 21; Length 5773;
 Best Local Similarity 100.0%; Pred. No. 8;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 2 catcgccagcatttc 16
 |||||
 Db 4886 CATGCCAGCATTTC 4872
 RESULT 6
 AAC74884
 ID AAC74884 standard; cDNA; 487 BP.

XX AAC74884;
 AC
 XX
 DT 08-FEB-2001 (first entry)
 XX
 DE Human ORFX ORF439 polynucleotide sequence SEQ ID NO:877.
 XX
 KW Human: open reading frame; ORFX; detection; cytostatic; hepatotropic;
 KW vulnary; antipsoriatic; antiparkinsonian; neurotropic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotrophic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
 KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KW thrombosis; contraceptive; ss.
 XX Homo sapiens.
 OS
 PN WO200058473-A2.
 XX
 PD 05-OCT-2000.
 XX
 PF 31-MAR-2000; 2000WO-US08621.
 XX
 PR 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 XX
 PA (CURA-) CURAGEN CORP.
 PI Shimkets RA, Leach M;
 XX
 DR WPI: 2000-602362/57.
 DR P-PSDB; AAB40675.
 XX
 PT Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 XX
 PS Claim 5; Page 899-900; 5507pp; English.
 XX
 CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vulnary;
 CC antipsoriatic; antiparkinsonian; neurotropic; neuroprotective;
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotrophic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
 CC antithyroid; and antianaemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.
 XX
 SQ Sequence 487 BP; 122 A; 99 C; 129 G; 137 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 487;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 cagcattgatcc 21
Db 316 cagcattgatcc 329
|||||

RESULT 7
AAC93854/c
ID AAC93854 standard; cDNA; 549 BP.

XX AC AAC93854;

XX DT 19-FEB-2001 (first entry)

XX DE Cat flea hindgut and Malpighian tubule (HMT) cDNA, SEQ ID NO:349.

XX KW Cat flea; hindgut and Malpighian tubule nucleic acid; HMT;
KW flea infestation; vaccine; antiparasitic; therapeutic target;
KW diagnosis; detection; ss.

XX OS Ctenocephalides felis.

XX PN W0200061621-A2.

XX PD 19-OCT-2000.

XX PF 07-APR-2000; 2000WO-US09437.

XX PR 09-APR-1999; 99US-0128704.

XX PA (HESK-) HESKA CORP.

XX PI Brandt KS, Gaines PJ, Stinchcomb DT, Wisniewski N;

XX WPI; 2000-656323/63.

XX PT Flea Malpighian tubule and head and nerve cord tissue derived nucleic
PT acids useful for the prevention, diagnosis and treatment of flea
PT infestations -

XX PS Claim 26; Page 361; 964pp; English.

XX CC The invention relates to novel cat flea (Ctenocephalides felis) nucleic
CC acids which are expressed in hindgut and Malpighian tubule (HMT) tissue
CC or head and nerve cord (HNC) tissue. The invention also relates to the
CC encoded proteins. The invention additionally encompasses expression
CC constructs, recombinant viruses and recombinant cells comprising the
CC nucleic acids of the invention, recombinant production of the proteins,
CC antibodies against the proteins, a method of identifying inhibitors of
CC the proteins, and compositions comprising the inhibitors for
CC administration to an animal. The nucleic acids, and the proteins they
CC encode may be used in the prevention, treatment and diagnosis of diseases
CC associated with flea infestations. For example, the nucleic acids may be
CC used to produce an HMT or HNC protein according to standard recombinant
CC DNA methodology by inserting the nucleic acids into a host cell and
CC culturing the cell to express the protein. The HMT and HNC nucleic acids
CC may also be used as DNA probes in diagnostic assays (e.g., PCR) to detect
CC and quantitate the presence of cat flea or other homologous nucleic acid
CC sequences in samples. They may also be used to study the expression and
CC function of the proteins and their role in metabolism. The HMT and HNC
CC proteins may be used as antigens in the production of specific
CC antibodies, and in assays to identify modulators (agonists and
CC antagonists) of HMT and/or HNC protein expression and activity. The
CC anti-HMT/HNC protein antibodies and antagonists may also be used to
CC downregulate protein expression and activity. The antibodies may also be
CC used as diagnostic agents for detecting the presence of flea polypeptides
CC in samples (e.g., by enzyme linked immunosorbent assay (ELISA)). The
CC present sequence represents a cat flea HMT cDNA of the invention.

XX SQ Sequence 549 BP; 155 A; 102 C; 130 G; 160 T; 2 other;

Query Match 56.0%; Score 14; DB 21; Length 549;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 catgccagcattt 15
Db 421 CATGCCAGCATT 408
|||||

RESULT 8
AAZ56381/c

ID AAZ56381 standard; DNA; 1344 BP.

XX AC AAZ56381;

XX DT 17-MAR-2000 (first entry)

XX DE Escherichia coli flagellin protein nucleotide sequence SEQ ID NO:51.

XX KW Flagellin; fliC; antigen; detection; ds.

XX OS Escherichia coli.

XX PN W09961458-A1.

XX PD 02-DEC-1999.

XX PF 21-MAY-1999; 99WO-AU00385.

XX PR 21-MAY-1998; 98AU-0003634.

XX PA (UNSY) UNIV SYDNEY.

XX PI Reeves PR, Wang L;

XX WPI; 2000-072598/06.

XX PT Novel nucleic acid molecule useful for the detection of flagellated
PT bacterial strains in food, faeces, etc. -

XX PS Claim 3; Page 225; 245pp; English.

XX CC AAZ56331 to AAZ56398 represent nucleic acid molecules (I) encoding all
CC or part of an Escherichia coli flagellin protein except a protein
CC expressed by E. coli H1, H7, H12 or H48 type strains. The present
CC invention also describes a method of detecting the presence of E. coli
CC of a particular H serotype in a sample, comprising specifically
CC hybridising a nucleic acid, preferably at least a pair, derived from a
CC flagellating gene, specific for a particular flagellin gene associated
CC with the H serotype, to any E. coli in the sample which contain the gene,
CC and detecting any hybridised molecules, identifying the presence of that
CC serotype in the sample. (I) are useful for: (1) detecting the presence
CC of E. coli of H serotype in a sample by hybridising at least one or a
CC pair of (I) to any E. coli in the sample and detecting the hybridised
CC nucleic acid molecules; and (2) for detecting the presence of both O
CC and H serotypes of E. coli by hybridising at least one or a pair of (I)
CC to any E. coli present in the sample and detecting the hybridised
CC nucleic acid molecules. (I) is particularly useful for detecting the
CC combination of O and H antigen. Hybridised (I) when using at least one
CC (I) is detected by southern blot analysis and, when using a pair of (I),
CC is detected by polymerase chain reaction (PCR). AAZ56399 to AAZ56420
CC represent primers used in the exemplification of the present invention.

XX SQ Sequence 1344 BP; 362 A; 337 C; 322 G; 323 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 1344;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcatgccagcatt 14


```

RESULT 11
AAV35235/c
ID AAV35235 standard; DNA; 5319 BP.
XX
XX AAV35235;
AC
XX
XX 28-SEP-1998 (first entry)
DT
XX
XX R. prowazekii S-layer protein genomic DNA.
DE
XX
XX Surface layer protein; S-layer; vaccine; typhus; spotted fever;
KW infection; diagnosis; disease; ds.
KW
XX
XX Rickettsia prowazekii.
OS
XX
XX Key Location/Qualifiers
PH -35_signal 340..345
FT /*tag= a
FT -10_signal 363..368
FT /*tag= b
FT RBS 379..386
FT /*tag= c
FT CDS 391..5229
FT /*tag= d
FT /product= S-layer protein
FT stem_loop 5270..5306
FT /*tag= e
XX
XX US5783441-A.
PN
XX
XX 21-JUL-1998.
XX
XX 20-DEC-1993; 93US-0169927.
XX
XX 20-DEC-1993; 93US-0169927.
PR
XX 09-AUG-1991; 91US-0742128.
PR
XX (USNA ) US SEC OF NAVY.
PA
XX
XX Carl M, Ching W, Dasch GA, Dobson MB;
PI
XX
XX WPI: 1998-427031/36.
DR
XX P-PSDB; AAW65088.
DR
XX
XX Recombinant DNA encoding Rickettsia surface layer proteins - useful
PT for diagnosing typhus and spotted fever and for preparing vaccines
PT against them
PT
XX
XX Claim 1; Column 11-24; 20pp; English.
PS
XX
XX This sequence encodes the Surface layer (S-layer) protein from
CC R. prowazekii strain Breinl. This sequence is useful for vaccination
CC against typhus and spotted fever rickettsial infection or for diagnosing
CC diseases caused by these bacteria. The surface layer protein antigens can
CC be produced recombinantly in large quantities.
XX
XX Sequence 5319 BP; 1815 A; 766 C; 983 G; 1755 T; 0 other;
SQ
Query Match 56.0%; Score 14; DB 19; Length 5319;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 catcgccagcattt 15
| | | | | | | | | | |
Db 2221 CATCGCCAGCAATT 2208
RESULT 12
AAAF21105/c
ID AAF21105 standard; DNA; 143068 BP.
XX
XX AAF21105;
AC

```

```

XX
DT 14-MAR-2001 (first entry)
DE
XX
XX Human low adenosine antisense oligonucleotide related sequence #2672.
XX
XX Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
KW human; airway disorder; bronchoconstriction; lung inflammation;
KW surfactant depletion; respiratory; bronchodilator; antiinflammatory;
KW immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;
KW respiratory obstruction; pulmonary vasoconstriction; impeded respiration;
KW surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
KW respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;
KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
KW cancer; ss.
XX
XX Homo sapiens.
OS
XX WO2000062736-A2.
PN
XX 26-OCT-2000.
PD
XX
XX 24-MAR-2000; 2000WO-US08020.
PF
XX
XX 06-APR-1999; 99US-0127958.
PR
XX (UYEC-) UNIV EAST CAROLINA.
XX (NYCE/) NYCE J W.
PA
XX
XX Nyce JW;
PI
XX
XX WPI: 2000-679539/66.
DR
XX
XX Low adenosine (A) content antisense oligonucleotides which do not
PT trigger adenosine receptors during metabolism, useful e.g. for treating
PT cancers and respiratory obstructions -
XX
XX Disclosure; Page 924-957; 1592pp; English.
PS
XX
XX The present invention describes low adenosine (A) content antisense
CC oligonucleotides and compositions (I) comprising them. In the antisense
CC oligonucleotides the A is replaced by a 'universal' or alternative base.
CC (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
CC immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
CC The antisense oligonucleotides and (I) can be used to down-regulate the
CC expression and or activity of target polypeptides associated with the
CC lung/respiratory disorders and malignancies, such as stimulating and
CC activating peptide factors and transmitters, transcription factors,
CC immunoglobulins and antibodies, antibody receptors, cytokines and
CC chemokines, endogenously produced specific and non-specific enzymes,
CC binding proteins, adhesion molecules and their receptors, cytokine and
CC chemokine receptors, adenosine receptors, bradykinin receptors, central
CC nervous system (CNS) and peripheral nervous and non-nervous system
CC receptors, CNS and peripheral nervous and non-nervous system peptide
CC transmitters, defensins, growth factors, vasoactive peptides and
CC receptors, binding proteins and malignancy associated proteins. The
CC antisense oligonucleotides may be used in this way to treat disorders
CC including respiratory obstruction (especially pulmonary obstruction
CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
CC and/or surfactant hypoproduction which are associated with a disease or
CC condition selected from pulmonary vasoconstriction, inflammation,
CC allergies, asthma, impeded respiration, respiratory distress syndrome
CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
CC pulmonary transplantation rejection, pulmonary infections, bronchitis,
CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
CC fragments and antisense oligonucleotides used in the exemplification of
CC the present invention.
XX
XX Sequence 143068 BP; 41194 A; 30122 C; 32403 G; 39349 T; 0 other;
SQ
Query Match 56.0%; Score 14; DB 21; Length 143068;

```

Best Local Similarity 100.0%; Pred. No. 26;
Matches 14; Conservative 0; Mismatches 0; Gaps 0;

QY 8 cagcatttcgattcc 21
|||||

Db 89829 CAGCATTTCGATCC 89816

RESULT 13
AAF21272/c
ID AAF21272 standard; DNA: 143068 BP.

XX AAF21272;
XX 14-MAR-2001 (first entry)

XX Human low adenosine antisense oligonucleotide related sequence #2839.

XX Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
KW human; airway disorder; bronchoconstriction; lung inflammation;
KW surfactant depletion; respiratory; bronchodilator; antiinflammatory;
KW immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;
KW respiratory obstruction; pulmonary obstruction; impeded respiration;
KW surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
KW respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
KW pulmonary hypertension; emphysema; pulmonary transplantation rejection;
KW chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
KW cancer; ss.

XX Homo sapiens.
OS
XX WO200062736-A2.
XX 26-OCT-2000.
XX 24-MAR-2000; 2000WO-US08020.
XX 06-APR-1999; 99US-0127958.
XX (UYEC-) UNIV EAST CAROLINA.
XX (NYCE/) NYCE J W.
XX Nyce JW;
XX WPI; 2000-679539/66.

XX Low adenosine (A) content antisense oligonucleotides which do not
PT trigger adenosine receptors during metabolism, useful e.g. for treating
PT cancers and respiratory obstructions -
XX
PS Disclosure: Page 1186-1219; 1592pp; English.

XX The present invention describes low adenosine (A) content antisense
CC oligonucleotides and compositions (I) comprising them. In the antisense
CC oligonucleotides the A is replaced by a 'Universal' or alternative base.
CC (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
CC immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
CC The antisense oligonucleotides and (I) can be used to down-regulate the
CC expression and/or activity of target polypeptides associated with
CC lung/respiratory disorders and malignancies, such as stimulating and
CC activating peptide factors and transmitters, transcription factors and
CC immunoglobulins and antibodies, antibody receptors, cytokines and
CC chemokines, endogenously produced specific and non-specific enzymes,
CC binding proteins, adhesion molecules and their receptors, cytokine and
CC chemokine receptors, adenosine receptors, bradykinin receptors, central
CC nervous system (CNS) and peripheral nervous and non-nervous system
CC receptors, CNS and peripheral nervous and non-nervous system peptide
CC transmitters, defensins, growth factors, vasoactive peptides and
CC receptors, binding proteins and malignancy associated proteins. The
CC antisense oligonucleotides may be used in this way to treat disorders
CC including respiratory obstruction (especially pulmonary obstruction
CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
CC and/or surfactant hypoproduction which are associated with a disease or

CC condition selected from pulmonary vasoconstriction, inflammation,
CC allergies, asthma, impeded respiration, respiratory distress syndrome
CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
CC pulmonary transplantation rejection, pulmonary infections, bronchitis,
CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
CC fragments and antisense oligonucleotides used in the exemplification of
CC the present invention.

XX Sequence 143068 BP; 41194 A; 30122 C; 32403 G; 39349 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 143068;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 cagcatttcgattcc 21
|||||

Db 89829 CAGCATTTCGATCC 89816

RESULT 14
AAA34983/c
ID AAA34983 standard; DNA: 143068 BP.

XX AAA34983;
XX 28-JUL-2000 (first entry)

XX Human adenosine receptor related polynucleotide SEQ ID NO:2672.

XX Human; adenosine receptor; low adenosine antisense oligonucleotide;
KW phosphorothioate; impaired respiration; inflammation; allergy;
KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
KW antiasthmatic; cytostatic; analgesic; impaired airway;
KW lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;
KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
KW cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.

XX Homo sapiens.
OS
XX WO200009525-A2.
XX 24-FEB-2000.
XX 03-AUG-1999; 99WO-US17712.
XX 03-AUG-1998; 98US-0095212.
XX (UYEC-) UNIV EAST CAROLINA.
XX Nyce JW;
XX WPI; 2000-205971/18.

XX New antisense oligonucleotides useful for treating e.g. pulmonary
PT vasoconstriction, inflammation, allergies, asthma, hypertension,
PT bronchitis, emphysema, respiratory distress syndrome, ischemia or
PT cancers -
XX
PS Disclosure: Page 851-882; 1343pp; English.

XX The present invention describes a new composition comprising an
CC antisense oligonucleotide (ON) with low adenosine (up to 15%), which
CC targets nucleic acids involved in bronchoconstriction, allergies, and/or
CC inflammation. The ON can have antiinflammatory, antiasthmatic,
CC antiasthmatic, cytostatic and analgesic activities. The compositions are
CC useful for the treatment of diseases associated with inflammation,
CC impaired airways, including lung disease and diseases whose secondary
CC effects afflict the lungs of a subject. They can be used for treating
CC e.g. ischaemic conditions, pulmonary vasoconstriction, allergies,
CC asthma, impeded respiration, respiratory distress syndrome, pain, cystic

CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive
 CC pulmonary disease (COPD), and cancers such as leukemias, lymphomas,
 CC carcinomas, and cancers which may metastasize to the lungs, including
 CC breast and prostate cancer. The reduction of the adenosine content of
 CC the ONS reduces side effects. The A-containing ONS break down with the
 CC release of deoxyadenosine which activates adenosine receptors causing
 CC bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
 CC nucleotide sequences given in the sequence listing from the present
 CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last
 CC 185 sequences are also called SEQ ID NO:1 to 185, but the sequences
 CC differ from the previously named sequences. SEQ ID NO:11 to 1680
 CC (AAA32323 to AAA33992) are specifically claimed ONS from the present
 CC invention. N.B. Sequences given in the disclosure of the present
 CC invention do not match up with their corresponding SEQ ID NO: sequences
 CC given in the sequence listing.

XX SQ Sequence 143068 BP; 41194 A; 30122 C; 32402 G; 39350 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 143068;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 cagcattcgatcc 21
 |||||

Db 89829 CAGCATTTCGATCC 89816

RESULT 15

AAA35150/c
 ID AAA35150 standard; DNA; 143068 BP.

XX AC AAA35150;

XX DT 28-JUL-2000 (first entry)

XX DE Human adenosine receptor related polynucleotide 2nd SEQ ID NO:24.

XX Human; adenosine receptor; low adenosine antisense oligonucleotide;
 KW phosphorothioate; impaired respiration; inflammation; allergy;
 KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
 KW antiallergic; antiasthmatic; cytotatic; analgesic; impaired airway;
 KW lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
 KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;
 KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
 KW cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.

XX OS Homo sapiens.

XX PN WO200009525-A2.

XX PD 24-FEB-2000.

XX PF 03-AUG-1999; 99WO-US17712.

XX PR 03-AUG-1998; 98US-0095212.

XX PA (UYEC-) UNIV EAST CAROLINA.

XX PI Nyce JW;

XX DR WPI; 2000-205971/18.

XX New antisense oligonucleotides useful for treating e.g. pulmonary
 PT vasoconstriction, inflammation, allergies, asthma, hypertension,
 PT bronchitis, emphysema, respiratory distress syndrome, ischemia or
 PT cancers

XX PS Disclosure; Page 1106-1138; 1343pp; English.

XX The present invention describes a new composition comprising an
 CC antisense oligonucleotide (ON) with low adenosine (up to 15%), which
 CC targets nucleic acids involved in bronchoconstriction, allergies, and/or

CC inflammation. The ON can have antiinflammatory, antiallergic,
 CC antiasthmatic, cytotatic and analgesic activities. The compositions are
 CC useful for the treatment of diseases associated with inflammation,
 CC impaired airways, including lung disease and diseases whose secondary
 CC effects afflict the lungs of a subject. They can be used for treating
 CC e.g. ischaemic conditions, pulmonary vasoconstriction, allergies,
 CC asthma, impaired respiration, respiratory distress syndrome, pain, cystic
 CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive
 CC pulmonary disease (COPD), and cancers such as leukemias, lymphomas,
 CC carcinomas, and cancers which may metastasize to the lungs, including
 CC breast and prostate cancer. The reduction of the adenosine content of
 CC the ONS reduces side effects. The A-containing ONS break down with the
 CC release of deoxyadenosine which activates adenosine receptors causing
 CC bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
 CC nucleotide sequences given in the sequence listing from the present
 CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last
 CC 185 sequences are also called SEQ ID NO:1 to 185, but the sequences
 CC differ from the previously named sequences. SEQ ID NO:11 to 1680
 CC (AAA32323 to AAA33992) are specifically claimed ONS from the present
 CC invention. N.B. Sequences given in the disclosure of the present
 CC invention do not match up with their corresponding SEQ ID NO: sequences
 CC given in the sequence listing.

XX SQ Sequence 143068 BP; 41194 A; 30126 C; 32402 G; 39346 T; 0 other;

Query Match 56.0%; Score 14; DB 21; Length 143068;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 cagcattcgatcc 21
 |||||

Db 89829 CAGCATTTCGATCC 89816

RESULT 16

AAA35151/c
 ID AAA35151 standard; DNA; 149412 BP.

XX AC AAA35151;

XX DT 28-JUL-2000 (first entry)

XX DE Human adenosine receptor related polynucleotide 2nd SEQ ID NO:25.

XX Human; adenosine receptor; low adenosine antisense oligonucleotide;
 KW phosphorothioate; impaired respiration; inflammation; allergy;
 KW allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
 KW antiallergic; antiasthmatic; cytotatic; analgesic; impaired airway;
 KW lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
 KW respiratory distress syndrome; pain; cystic fibrosis; emphysema;
 KW pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
 KW cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.

XX OS Homo sapiens.

XX PN WO200009525-A2.

XX PD 24-FEB-2000.

XX PF 03-AUG-1999; 99WO-US17712.

XX PR 03-AUG-1998; 98US-0095212.

XX PA (UYEC-) UNIV EAST CAROLINA.

XX PI Nyce JW;

XX DR WPI; 2000-205971/18.

XX New antisense oligonucleotides useful for treating e.g. pulmonary
 PT vasoconstriction, inflammation, allergies, asthma, hypertension,
 PT bronchitis, emphysema, respiratory distress syndrome, ischemia or

PT cancers -
XX Disclosure; Page 1138-1171; 1343pp; English.
XX

CC The present invention describes a new composition comprising an
CC antisense oligonucleotide (ON) with low adenosine (up to 15%), which
CC targets nucleic acids involved in bronchoconstriction, allergies, and/or
CC inflammation. The ON can have antiinflammatory, antiallergic,
CC antiasthmatic, cytostatic and analgesic activities. The compositions are
CC useful for the treatment of diseases associated with inflammation,
CC impaired airways, including lung disease and diseases whose secondary
CC effects afflict the lungs of a subject. They can be used for treating
CC e.g. ischaemic conditions, pulmonary vasoconstriction, allergies,
CC asthma, impeded respiration, respiratory distress syndrome, pain, cystic
CC fibrosis, pulmonary hypertension, emphysema, chronic obstructive
CC pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,
CC carcinomas, and cancers which may metastasize to the lungs, including
CC breast and prostate cancer. The reduction of the adenosine content of
CC the ONs reduces side effects. The A-containing ONs break down with the
CC release of deoxyadenosine which activates adenosine receptors causing
CC bronchoconstriction and inflammation. AAA32313 to AAA35312 represent the
CC nucleotide sequences given in the sequence listing from the present
CC invention, which correspond to SEQ ID NO:1 to 2815, and then the last
CC 185 sequences are also called SEQ ID NO:1 to 185, but the sequences
CC differ from the previously named sequences. SEQ ID NO:11 to 1680
CC (AAA32323 to AAA33992) are specifically claimed ONs from the present
CC invention. N.B. Sequences given in the disclosure of the present
CC invention do not match up with their corresponding SEQ ID NO: sequences
CC given in the sequence listing.
XX

XX Sequence 149412 BP; 43049 A; 31388 C; 33852 G; 41123 T; 0 other;
SQ

Query Match 56.0%; Score 14; DB 21; Length 149412;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 cagcatttcgatcc 21
|||||
Db 96173 CAGCATTTCGATCC 96160

RESULT 17
AAF21273/c
ID AAF21273 standard; DNA; 152740 BP.
XX
AC AAF21273;
XX
DT 14-MAR-2001 (first entry)
XX
XX Human low adenosine antisense oligonucleotide related sequence #2840.
XX

Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
human; airway disorder; bronchoconstriction; lung inflammation;
surfactant depletion; respiratory; bronchodilator; antiinflammatory;
immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;
respiratory obstruction; pulmonary obstruction; impeded respiration;
surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
pulmonary hypertension; emphysema; pulmonary transplantation rejection;
chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
cancer; ss.
XX
XX Homo sapiens.
OS
XX WO2000062736-A2.
PN
XX 26-OCT-2000.
PD
XX 24-MAR-2000; 2000WO-US08020.
PF
XX 06-APR-1999; 99US-0127958.
PR
XX

(UYEC-) UNIV EAST CAROLINA.
(NYCE/) NYCE J W.
PI Nyce JW;
XX
XX WPI; 2000-679539/66.
XX

Low adenosine (A) content antisense oligonucleotides which do not
trigger adenosine receptors during metabolism, useful e.g. for treating
cancers and respiratory obstructions -
XX
XX Disclosure; Page 1219-1254; 1592pp; English.
XX

CC The present invention describes low adenosine (A) content antisense
CC oligonucleotides and compositions (I) comprising them. In the antisense
CC oligonucleotides the A is replaced by a 'Universal' or alternative base.
CC (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
CC immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
CC The antisense oligonucleotides and (I) can be used to down-regulate the
CC expression and or activity of target polypeptides associated with the
CC lung/respiratory disorders and malignancies, such as stimulating and
CC activating peptide factors and transmitters, transcription factors,
CC immunoglobulins and antibodies, antibody receptors, cytokines and
CC chemokines, endogenously produced specific and non-specific enzymes,
CC binding proteins, adhesion molecules and their receptors, cytokine and
CC chemokine receptors, adenosine receptors, bradykinin receptors, central
CC nervous system (CNS) and peripheral nervous and non-nervous system
CC receptors, CNS and peripheral nervous and non-nervous system peptide
CC transmitters, defensins, growth factors, vasoactive peptides and
CC receptors, binding proteins and malignancy associated proteins. The
CC antisense oligonucleotides may be used in this way to treat disorders
CC including respiratory obstruction (especially pulmonary obstruction
CC and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
CC and/or surfactant hypoproduction which are associated with a disease or
CC condition selected from pulmonary vasoconstriction, inflammation,
CC allergies, asthma, impeded respiration, respiratory distress syndrome
CC (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
CC hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
CC pulmonary transplantation rejection, pulmonary infections, bronchitis,
CC and/or cancer. AAF18434 to AAF21543 represent human polynucleotide
CC fragments and antisense oligonucleotides used in the exemplification of
CC the present invention.
XX

XX Sequence 152740 BP; 44169 A; 32023 C; 34549 G; 41999 T; 0 other;
SQ

Query Match 56.0%; Score 14; DB 21; Length 152740;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 cagcatttcgatcc 21
|||||
Db 96173 CAGCATTTCGATCC 96160

RESULT 18
AAV11150/c
ID AAV11150 standard; DNA; 22 BP.
XX
XX AAV11150;
AC
XX
XX 14-JUL-1998 (first entry)
DT
XX
XX Oligonucleotide #3 for nucleic acid assay.
DE
XX
XX Nucleic acid assay; hybridising; probe; target; duplex; detection; ss.
XX
XX Synthetic.
OS
XX DE19633436-A1.
PN
XX 26-FEB-1998.
PD
XX

PF 20-AUG-1996; 96DE-1033436.
XX 20-AUG-1996; 96DE-1033436.
PR (BOEF) BOEHRINGER MANNHEIM GMBH.
XX Bergmann F, Herrmann R, Kobold U;
XX WPI; 1998-146248/14.
XX Nucleic acid assay - based on mass of probe-target duplex
XX Disclosure; Fig 1; 9pp; German.
XX AAV11148-V11151 are oligonucleotides used in a novel nucleic acid assay.
CC This assay involves hybridising a probe to a target sequence of the
CC nucleic acid to form a duplex and degrading the nucleic acid to form
CC a fragment of a defined length containing the duplex, which is detected
CC on the basis of its mass. This method enables different nucleic acids to
CC be detected simultaneously does not require the use of labelled probes.
XX Sequence 22 BP; 5 A; 3 C; 7 G; 7 T; 0 other;
SQ

Query Match 52.0%; Score 13; DB 19; Length 22;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 catcgccagcatt 14
Db 18 CATCGCCAGCATT 6

RESULT 19
AAA76686
ID AAA76686 standard; cDNA; 51 BP.
AC AAA76686;
XX 16-NOV-2000 (first entry)
DT Human clone cg28389525 polymorphic site, SEQ ID NO:369.
DE Human; single nucleotide polymorphism: SNP;
XX detection; identification; gene therapy; ss.
KW Homo sapiens.
OS
XX
XX Key Location/Qualifiers
FH variation replace (26,C) /*tag= a
FT
FT
XX WO200029623-A2.
PN 25-MAY-2000.
XX 17-NOV-1999; 99WO-US27293.
XX 17-NOV-1998; 98US-0109024.
XX 16-NOV-1999; 99US-0109024.
XX (CURA-) CURAGEN CORP.
XX Shimkets RA, Leach MD;
XX WPI; 2000-387826/33.
XX Human nucleic acids containing single nucleotide polymorphisms, useful
XX for treating a subject suffering, or at risk from a pathology due to
XX the presence of a sequence polymorphism -
XX Claim 1; Page 269; 543pp; English.
PS
PS

CC Sequences AAA76318-A77509 represent 1192 human nucleic acid sequences
CC which contain single nucleotide polymorphisms (SNPs). Sequences 1 to
CC 1112 (AAA76318-A77429) are consecutive pairs of nucleotides which
CC contain silent SNPs. Sequences 1113 to 1192 (AAA77430-A77509) are
CC consecutive pairs of nucleotides containing SNPs which result in changes
CC in the corresponding amino acid sequences (AAB11749-B11828). The SNPs in
CC sequences 1113 to 1128 (AAA77430-A77445) lead to conservative amino acid
CC changes, while those in sequences 1129 to 1186 (AAA77446-A77503) result
CC in non-conservative changes. The SNPs in sequences 1187 to 1192
CC (AAA77504-A77509) generate frameshift mutations. The invention also
CC relates to a method of detecting a polymorphic site in a nucleic acid and
CC a method of determining the relatedness of two nucleic acids. It also
CC encompasses peptides containing polymorphic sites, antibodies raised
CC against such peptides, and a method of detecting polymorphic
CC proteins/peptides using the antibodies. The nucleic acids are useful for
CC gene therapy of an individual having, suspected of having, or at risk of
CC developing a pathological condition due to the presence of a sequence
CC polymorphism. Such treatment would comprise administration of the
CC wild-type nucleic acid sequence. Antibodies raised against polymorphic
CC peptides can also be used in the treatment of such individuals.
XX Sequence 51 BP; 13 A; 11 C; 15 G; 12 T; 0 other;
SQ

Query Match 52.0%; Score 13; DB 21; Length 51;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 5 cgccagcattcg 17
Db 11 cgccagcattcg 23

RESULT 20
AAA76687
ID AAA76687 standard; cDNA; 51 BP.
AC AAA76687;
XX 16-NOV-2000 (first entry)
DT Human clone cg28389525 polymorphic site, SEQ ID NO:370.
DE Human; single nucleotide polymorphism: SNP;
XX detection; identification; gene therapy; ss.
KW Homo sapiens.
OS
XX
XX Key Location/Qualifiers
FH variation replace (26,T) /*tag= a
FT
FT
XX WO200029623-A2.
PN 25-MAY-2000.
XX 17-NOV-1999; 99WO-US27293.
XX 17-NOV-1998; 98US-0109024.
XX 16-NOV-1999; 99US-0109024.
XX (CURA-) CURAGEN CORP.
XX Shimkets RA, Leach MD;
XX WPI; 2000-387826/33.
XX Human nucleic acids containing single nucleotide polymorphisms, useful
XX for treating a subject suffering, or at risk from a pathology due to
XX the presence of a sequence polymorphism -
XX Claim 1; Page 270; 543pp; English.
PS
PS

Sequences AAA76318-A77509 represent 1192 human nucleic acid sequences which contain single nucleotide polymorphisms (SNPs). Sequences 1 to 1112 (AAA76318-A77429) are consecutive pairs of nucleotides which contain silent SNPs. Sequences 1113 to 1192 (AAA77430-A77509) are consecutive pairs of nucleotides containing SNPs which result in changes in the corresponding amino acid sequences (AAB11749-B11828). The SNPs in sequences 1113 to 1128 (AAA77430-A77445) lead to conservative amino acid changes, while those in sequences 1129 to 1186 (AAA77446-A77503) result in non-conservative changes. The SNPs in sequences 1187 to 1192 (AAA77504-A77509) generate frameshift mutations. The invention also relates to a method of detecting a polymorphic site in a nucleic acid and a method of determining the relatedness of two nucleic acids. It also encompasses peptides containing polymorphic sites, antibodies raised against such peptides, and a method of detecting polymorphic proteins/peptides using the antibodies. The nucleic acids are useful for gene therapy of an individual having, suspected of having, or at risk of developing a pathological condition due to the presence of a sequence polymorphism. Such treatment would comprise administration of the wild-type nucleic acid sequence. Antibodies raised against polymorphic peptides can also be used in the treatment of such individuals.

Sequence 51 BP; 13 A; 12 C; 15 G; 11 T; 0 other;

Query Match 52.0%; Score 13; DB 21; Length 51;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 cgcgcagcttcg 17
|||||

Db 11 cgcgcagcttcg 23

RESULT 21

AAZ19606
ID AAZ19606 standard; RNA; 61 BP.

AC AAZ19606;

DT 08-NOV-1999 (first entry)

DE Complement system protein C5 RNA binding ligand 13.

XX Complement system protein; Clq; C5; C3; RNA ligand; neurotropic; stroke;
KW neuroprotective; anti-asthmatic; cerebroprotective; immunosuppressive;
KW diagnostic; prevention; treatment; complement protein-related disease;
KW Alzheimer's disease; renal disease; transplant rejection; asthma; ss.

OS Synthetic.

XX WO9941271-A1.

XX 19-AUG-1999.

XX 05-FEB-1999; 99WO-US02597.

XX 29-SEP-1998; 98US-0163025.

XX 12-FEB-1998; 98US-0023228.

PA (NEXS-) NEXSTAR PHARM INC.

XX Biesecker G, Gold L;

XX WPI; 1999-527357/44.

XX New Nucleic Acid Ligand to complement protein C5, useful in the
PT treatment of Alzheimer's disease, renal diseases, transplant
PT rejection, stroke and asthma

PS Example 4; Page 42; 120pp; English.

XX This invention describes novel purified and isolated non-naturally
CC occurring RNA ligands (I) to complement system proteins C5, C1q and C3.

CC The products of the invention have neurotropic, neuroprotective,
CC anti-asthmatic, cerebroprotective and immunosuppressive activity. The
CC new ligands are useful as diagnostic agents, and pharmaceutical agents
CC for prevention and treatment of complement protein-related diseases,
CC including Alzheimer's disease, renal diseases, transplant rejection,
CC stroke and asthma.

XX Sequence 61 BP; 11 A; 18 C; 21 G; 11 U; 0 other;

Query Match 52.0%; Score 13; DB 20; Length 61;
Best Local Similarity 69.2%; Pred. No. 1.2e+02;
Matches 9; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 8 cagcatttcgac 20

Db 23 cagcauuucgac 35

RESULT 22

AAZ14744/C
ID AAZ14744 standard; cDNA; 300 BP.

XX AAZ14744;

XX 12-OCT-1999 (first entry)

XX Human gene expression product cDNA sequence SEQ ID NO:2213.

XX Human; gene; gene expression product; diagnosis; therapy; probe;
KW detection; mapping; tissue typing; profiling; forensic; cancer;
KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.

OS Homo sapiens.

XX WO9938972-A2.

XX 05-AUG-1999.

XX 28-JAN-1999; 99WO-US01619.

XX 03-APR-1998; 98US-0080666.

XX 28-JAN-1998; 98US-0072910.

XX 24-FEB-1998; 98US-0075954.

XX 31-MAR-1998; 98US-0080114.

XX 03-APR-1998; 98US-0080515.

XX (CHIR) CHIRON CORP.

XX (HYSE-) HYSEQ INC.

XX Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;
PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;
PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;
PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
PI Stache-Crain B, Sudduth-Klinger J, Williams LT;

XX WPI; 1999-494092/41.

XX Novel human genes and their expression products which are
PT differentially expressed in different cell types

XX Claim 1; Page 1159; 2479pp; English.

XX The present invention describes a library of human polynucleotides
CC comprising the sequences given in AAZ12532 to AAZ17779. Also described is
CC a method of detecting differentially expressed genes correlated with the
CC cancerous state of a mammalian cell, comprising detecting at least one
CC differentially expressed gene product in a test sample from a cell
CC suspected of being cancerous, where the gene product is encoded by one
CC of the 5248 polynucleotide sequences given in AAZ12532 to AAZ17779. The
CC polynucleotides can be used as a source of primers and probes, which can
CC be used for a variety of purpose, e.g. detection of expression levels,
CC mapping, tissue typing or profiling, forensics, genetic analysis and

CC detection of polymorphisms. Polypeptides encoded by the polynucleotides
 CC can be used for raising antibodies for experimental, diagnostic and
 CC therapeutic purposes. The polynucleotides may also be used to construct
 CC arrays for diagnostics (which may be used to determine function of an
 CC encoded protein); and to detect differences in expression levels between
 CC two cells (e.g. to identify abnormal or diseased tissue in a human, to
 CC identify a genetic predisposition or susceptibility to a disease such as
 CC cancer). The polynucleotides of the invention are especially used in the
 CC diagnosis, prognosis and management of colorectal cancer, breast cancer,
 CC and lung cancer. The polynucleotides can also be used to screen for
 CC peptide analogues and antagonists.

XX SQ Sequence 300 BP; 83 A; 62 C; 70 G; 85 T; 0 other;

Query Match 52.0%; Score 13; DB 20; Length 300;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gcatgccagcat 13
 |||||
 Db 174 GCATGCCAGCAT 162

RESULT 23.
 AAA31174/c

ID AAA31174 standard; DNA; 322 BP.

XX AC AAA31174;

XX DT 05-JUL-2000 (first entry)

XX DE Plant microsatellite marker #135.

XX KW Plant microsatellite sequence; core repeat sequence; detection; probe;
 KW DNA polymorphism; genome mapping; physical mapping; fingerprinting;
 KW variety identification; genetic variability evaluation; primer; ss.

XX OS Eucalyptus grandis.

XX PN W09967421-A1.

XX PD 29-DEC-1999.

XX PF 25-JUN-1999; 99WO-NZ00092.

XX PR 25-JUN-1998; 98US-0105307.

XX PA (GENE-) GENESIS RES & DEV CORP LTD & FLETCHER.
 PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.

XX PI Havukkala IJ, Bloksberg LN, Glenn M;

XX DR WPI; 2000-116958/10.

XX PT New plant microsatellite markers and associated flanking species for
 PT the detection of polymorphic genetic markers -

XX PS Claim 1; Page 116; 392pp; English.

XX CC Sequences AAA31040-A32093 represent novel plant microsatellite sequences
 CC and associated flanking species. The sequences comprise a central core
 CC repeat sequence, especially selected from the sequences AAA32094-A32096
 CC with left and right flanking sequences. The polynucleotide sequences
 CC can be used in the detection of DNA polymorphisms, in genome mapping,
 CC in physical mapping, in positional cloning of genes, in variety
 CC identification and in evaluation of genetic variability within and
 CC between plant tissues, populations, cultivars, species and species
 CC groups. They may also be used to design hybridization probes for
 CC oligonucleotide fingerprinting and library screening and to design
 CC primers for microsatellite-primed PCR. Microsatellite markers are
 CC useful to locate specific economically useful genes in plant genomes.

SQ Sequence 322 BP; 43 A; 101 C; 88 G; 89 T; 1 other;

Query Match 52.0%; Score 13; DB 21; Length 322;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 atcgccagcat 15
 |||||
 Db 21 ATCGCCAGCAT 9

RESULT 24

AAT27962

ID AAT27962 standard; DNA; 447 BP.

XX AC AAT27962;

XX DT 11-MAR-1997 (first entry)

XX DE Hepatitis C virus type 10a isolate NN58 bases 478-925.

XX KW Hepatitis C virus; subtype; polymerase chain reaction; amplification;
 KW PCR; primer; probe; antibody; infection; ss.

XX OS Hepatitis C virus.

XX PN W09613590-A2.

XX PD 09-MAY-1996.

XX PF 23-OCT-1995; 95WO-RP04155.

XX PR 28-JUN-1995; 95EP-0870076.

XX PR 21-OCT-1994; 94EP-0870166.

XX PA (INNO-) INNOGENETICS NV.

XX PI Maertens G, Stuyver L;

XX DR WPI; 1996-251460/25.

XX DR P-PSDB; AAR96551.

XX PT Hepatitis C virus polynucleic acid unique to unidentified sub: type
 PT - used to develop probes and primers for new sub: types and vaccines
 PT to prevent and treat infection

XX PS Claim 6; Fig 3; 150pp; English.

XX CC The sequences AAT27937-T27989 represent novel sequences isolated from
 CC hepatitis C virus subtypes different from subtypes 1a-c, 2a-d, 3a-f,
 CC 4a-j, 5a and 6a. They esp. from the novel subtypes 1d-f, 2e-i, 2k, 2l,
 CC 3g, 4k-m, 7a-c or types 9, 10 or 11. The sequences corresp. to the 5',
 CC untranslated region (UR), the Core/E1, NS4 or NS5B regions of the
 CC genome. This sequence represents nucleotides 478-925 from the HCV type
 CC 10a isolate NE98.

XX CC The new HCV types were isolated from patients with chronic HCV from the
 CC Benelux countries, France, Cameroon and Vietnam, because of their
 CC aberrant reactivities. The RNA was extracted, cDNA synthesised and PCR
 CC amplified, cloned and genotyped. The 5' UR, Core/E1 and NS5B regions were
 CC sequenced either directly or partially and used to classify the new
 CC viruses into (sub)types based on comparison with known sequences.
 CC The sequences were used to generate the peptides AAR96424-R96524. The
 CC sequences can also be used to synthesise probes and primers for the
 CC detection of HCV in a sample. The polypeptides can be used to detect
 CC anti-HCV antibodies, for HCV typing or to prevent HCV infections.

XX SQ Sequence 447 BP; 82 A; 130 C; 114 G; 118 T; 3 other;

Query Match 52.0%; Score 13; DB 17; Length 447;

Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 catgccagcatt 14
Db 406 catgccagcatt 418
RESULT 25
AAC54077
ID AAC54077 standard; DNA; 591 BP.
XX AC AAC54077;
XX DT 18-OCT-2000 (first entry)
XX DE Arabidopsis thaliana DNA fragment SEQ ID NO: 76608.
XX KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 07-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 19-MAY-1999; 99US-0134768.
PR 20-MAY-1999; 99US-0134941.
PR 21-MAY-1999; 99US-0135124.
PR 24-MAY-1999; 99US-0135353.
PR 25-MAY-1999; 99US-0135629.
PR 27-MAY-1999; 99US-0136021.
PR 28-MAY-1999; 99US-0136392.
PR 01-JUN-1999; 99US-0136782.
PR 03-JUN-1999; 99US-0137222.
PR 04-JUN-1999; 99US-0137528.
PR 07-JUN-1999; 99US-0137502.
PR 08-JUN-1999; 99US-0137724.
PR 10-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0140991.
PR 01-JUL-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.

PR 13-AUG-1999; 99US-0148684.
 PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149175.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149902.
 PR 23-AUG-1999; 99US-0149930.
 PR 25-AUG-1999; 99US-0150566.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 27-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157553.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 52.0%; Score 13; DB 21; Length 591;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 attcgatctctcg 24

Db 144 attcgatctctcg 156

RESULT 26

AAC35502
 ID AAC35502 standard; DNA; 606 BP.
 XX
 AC AAC35502;
 XX
 DT 17-OCT-2000 (first entry)
 XX
 DE Arabidopsis thaliana DNA fragment SEQ ID NO: 10429.
 XX
 KW Hybridisation assay; genetic mapping; gene expression control;
 KW protein identification; signal transduction pathway;
 KW metabolic pathway; promoter; termination sequence; ss.
 XX
 OS Arabidopsis thaliana.
 XX
 PN EP1033405-A2.
 XX
 PD 06-SEP-2000.
 XX
 PF 25-FEB-2000; 2000EP-0301439.
 XX
 PR 25-FEB-1999; 99US-0121825.
 PR 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126264.
 PR 29-MAR-1999; 99US-0126785.
 PR 01-APR-1999; 99US-0127462.
 PR 06-APR-1999; 99US-0128234.
 PR 08-APR-1999; 99US-0128714.
 PR 16-APR-1999; 99US-0129845.
 PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 23-APR-1999; 99US-0130891.
 PR 28-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132048.
 PR 30-APR-1999; 99US-0132407.
 PR 04-MAY-1999; 99US-0132484.
 PR 05-MAY-1999; 99US-0132485.
 PR 06-MAY-1999; 99US-0132486.
 PR 07-MAY-1999; 99US-0132487.
 PR 11-MAY-1999; 99US-0132863.
 PR 14-MAY-1999; 99US-0134256.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134221.
 PR 14-MAY-1999; 99US-0134370.
 PR 18-MAY-1999; 99US-0134768.
 PR 19-MAY-1999; 99US-0134941.
 PR 20-MAY-1999; 99US-0135124.
 PR 21-MAY-1999; 99US-0135353.
 PR 24-MAY-1999; 99US-0135629.
 PR 25-MAY-1999; 99US-0136021.
 PR 27-MAY-1999; 99US-0136392.
 PR 28-MAY-1999; 99US-0136782.
 PR 01-JUN-1999; 99US-0137222.
 PR 03-JUN-1999; 99US-0137528.
 PR 04-JUN-1999; 99US-0137502.
 PR 07-JUN-1999; 99US-0137724.
 PR 08-JUN-1999; 99US-0138094.
 PR 10-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139452.
 PR 16-JUN-1999; 99US-0139453.
 PR 17-JUN-1999; 99US-0139492.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.

PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 24-JUN-1999; 99US-0140823.
PR 28-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145088.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.

PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 26-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 52.0%; Score 13; DB 21; Length 606;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 attcgatctctcg 24
Db 159 attcgatctctcg 171

RESULT 27
AAF09259
ID AAF09259 standard; cDNA; 629 BP.
XX
AC AAF09259;
XX
DT 13-MAR-2001 (first entry)
XX

DE Fusarium venenatum EST SEQ ID NO:1782.

XX Multiple gene expression; filamentous fungal cell; EST;

XX expressed sequence tag; Fusarium venenatum; Aspergillus niger;

KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;

KW culture condition; environmental stress; spore morphogenesis;

KW metabolic pathway engineering; catabolic pathway engineering; ss.

XX Fusarium venenatum.

OS

XX

XX W0200056762-A2.

PN

XX

XX 28-SEP-2000.

PD

XX

XX 22-MAR-2000; 2000WO-US07781.

PF

XX

XX 22-MAR-1999; 99US-0273623.

PR

XX (NOVO) NOVO NORDISK BIOTECH INC.

PA

XX (NOVO) NOVO NORDISK AS.

PA

XX Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;

PI

XX WPI; 2000-594572/56.

DR

XX

XX Monitoring differential expression of genes in filamentous fungal cells

CC uses fluorescence-labeled nucleic acids isolated from the cells and a

CC substrate of expressed sequence tags -

CC

XX Claim 86; Page 1051; 3161pp; English.

PS

XX

CC The present invention describes a method for monitoring differential

CC expression of genes in a first filamentous fungal (FF) cell relative to

CC expression of the same genes in one or more second filamentous fungal

CC cells. The method uses fluorescence-labeled nucleic acids isolated from

CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs

CC are used in the methods for monitoring differential expression of genes

CC in a first filamentous fungal (FF) cell relative to expression of the

CC same genes in one or more second filamentous fungal cells. Monitoring

CC the global expression of genes from FF cells allows the production

CC potential of the microorganisms to be improved. New genes may be

CC discovered, possible functions of unknown open reading frames can be

CC identified and gene copy number variation and stability can be

CC monitored. The expression of genes can be used to study how FF cells

CC adapt to changes in culture conditions, environmental stress, spore

CC morphogenesis, recombination, metabolic or catabolic pathway

CC engineering. Using ESTs provides several advantages over genomic or

CC random cDNA clones including elimination of redundancy as one spot on an

CC array equals one gene or open reading frame, and organisation of the

CC microarrays based on function of the gene products to facilitate

CC analysis of the results. AAF07478 to AAF11247 represents ESTs from

CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus

CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and

CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are

CC all specifically claimed in the present invention.

XX

XX Sequence 629 BP; 153 A; 169 C; 135 G; 158 T; 4 other;

SQ

Query Match 52.0%; Score 13; DB 21; Length 629;

Best Local Similarity 100.0%; Pred. No. 1.1e+02;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 catcgccagcatt 14

DB 83 catcgccagcatt 95

|||||

RESULT 28

AAZ17005

ID AAZ17005 standard; cDNA; 714 BP.

XX

AC AAZ17005;

XX 12-OCT-1999 (first entry)

XX

XX Human gene expression product cDNA sequence SEQ ID NO:4475.

DE

XX

XX Human; gene: gene expression product; diagnosis; therapy; probe;

KW detection; mapping; tissue typing; profiling; forensic; cancer;

KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.

XX

XX Homo sapiens.

OS

XX

XX W09938972-A2.

PN

XX

XX 05-AUG-1999.

PD

XX

XX 28-JAN-1999; 99WO-US01619.

PF

XX

XX 03-APR-1998; 98US-0080666.

PR

XX 28-JAN-1998; 98US-0072910.

PR

XX 24-FEB-1998; 98US-0075954.

PR

XX 31-MAR-1998; 98US-0080114.

PR

XX 03-APR-1998; 98US-0080515.

PR

XX (CHIR) CHIRON CORP.

PA

XX (HYSE) HYSEQ INC.

PA

XX Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;

PI Escobedo J, Garcia PB, Garcia V, Giese K, Innis MA;

PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;

PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;

PI Stache-Crain B, Sudduth-Klinger J, Williams LT;

PI

XX WPI; 1999-494092/41.

DR

XX

XX Novel human genes and their expression products which are

PT differentially expressed in different cell types

PT

XX Claim 1; Page 2120; 2479pp; English.

PS

XX

XX The present invention describes a library of human polynucleotides

CC comprising the sequences given in AAZ12532 to AAZ17779. Also described is

CC a method of detecting differentially expressed genes correlated with the

CC cancerous state of a mammalian cell, comprising detecting at least one

CC differentially expressed gene product in a test sample from a cell

CC suspected of being cancerous, where the gene product is encoded by one

CC of the 5248 polynucleotide sequences given in AAZ12532 to AAZ17779. The

CC polynucleotides can be used as a source of primers and probes, which can

CC be used for a variety of purpose, e.g. detection of expression levels,

CC mapping, tissue typing or profiling, forensics, genetic analysis and

CC detection of polymorphisms. Polypeptides encoded by the polynucleotides

CC can be used for raising antibodies for experimental, diagnostic and

CC therapeutic purposes. The polynucleotides may also be used to construct

CC arrays for diagnostics (which may be used to determine function of an

CC encoded protein); and to detect differences in expression levels between

CC two cells (e.g. to identify abnormal or diseased tissue in a human, to

CC identify a genetic predisposition or susceptibility to a disease such as

CC cancer). The polynucleotides of the invention are especially used in the

CC diagnosis, prognosis and management of colorectal cancer, breast cancer,

CC and lung cancer. The polynucleotides can also be used to screen for

CC peptide analogues and antagonists.

XX

XX Sequence 714 BP; 186 A; 142 C; 146 G; 231 T; 9 other;

SQ

Query Match 52.0%; Score 13; DB 20; Length 714;

Best Local Similarity 100.0%; Pred. No. 1.1e+02;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 tcgcccagcattc 16

DB 74 tcgcccagcattc 86

|||||

RESULT 29
AAF22436
ID AAF22436 standard; cDNA; 725 BP.
XX
XX AAF22436;
AC
DT 26-MAR-2001 (first entry)
XX
XX Human breast cancer associated antigen nucleotide sequence SEQ ID NO:15.
DE
XX Human; breast cancer; gastric cancer; prostate cancer; diagnosis;
KW cancer associated antigen; cytostatic; cancer vaccine; ss.
XX
XX Homo sapiens.
OS
XX WO200073801-A2.
PN
XX 07-DEC-2000.
PD
XX
XX 26-MAY-2000; 2000WO-US14749.
PF
XX
PR 28-MAY-1999; 99US-0136526.
PR 10-SEP-1999; 99US-0153454.
XX
XX (LUDW-) LUDWIG INST CANCER RES.
PA
XX
XX Obata Y;
PI
XX WPI; 2001-025274/03.
DR
XX
XX Nucleic acids encoding breast, gastric and prostate cancer associated
PT antigen precursors, useful for diagnosing and treating a condition
PT characterized by expression of an abnormal amount of a protein, e.g.
PT cancer -
XX
XX Claim 50; Page 276; 799pp; English.
PS
XX AAF22422 to AAF22626, AAF22627 to AAF22773 and AAF22774 to AAF23014
XX represent nucleotide sequences encoding human breast, gastric and
CC prostate cancer associated antigen precursors (CAAP) respectively.
CC AAB63232 to AAB63467, AAB63468 to AAB63721 and AAB63722 to AAB63970
CC represent human breast, gastric and prostate CAAP protein sequence
CC respectively. CAAPs have cytostatic activity and can be used in the
CC production of cancer vaccines. The human CAAP proteins, peptides, nucleic
CC acids or anti-CAAP antibodies are useful for diagnosing and treating a
CC condition characterised by expression of an abnormal amount of a protein,
CC e.g. cancer.
XX
XX Sequence 725 BP; 131 A; 225 C; 224 G; 136 T; 9 other;
SQ

Query Match 52.0%; Score 13; DB 22; Length 725;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gcatgccagcat 13
Db 229 gcatgccagcat 241
|||||

RESULT 30
AAZ17191/C
ID AAZ17191 standard; cDNA; 750 BP.
XX
XX AAZ17191;
AC
XX
XX 12-OCT-1999 (first entry)
DT
XX Human gene expression product cDNA sequence SEQ ID NO:4662.
DE
XX Human; gene; gene expression product; diagnosis; therapy; probe;
KW detection; mapping; tissue typing; profiling; forensic; cancer;
KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.

XX Homo sapiens.
OS
XX WO9938972-A2.
PN
XX 05-AUG-1999.
PD
XX
XX 28-JAN-1999; 99WO-US01619.
PF
XX
PR 03-APR-1998; 98US-0080666.
PR 28-JAN-1998; 98US-0072910.
PR 31-FEB-1998; 98US-0075954.
PR 24-MAR-1998; 98US-0080114.
PR 03-APR-1998; 98US-0080515.
XX
XX (CHIR) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX
XX Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;
PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;
PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;
PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
PI Stache-Crain B, Sudduth-Klinger J, Williams LT;
XX
XX WPI; 1999-494092/41.
DR
XX
XX Novel human genes and their expression products which are
PT differentially expressed in different cell types
PT
XX
XX Claim 1; Page 2213-2214; 2479pp; English.
PS
XX The present invention describes a library of human polynucleotides
CC comprising the sequences given in AAZ12532 to AAZ1779. Also described is
CC a method of detecting differentially expressed genes correlated with the
CC cancerous state of a mammalian cell, comprising detecting at least one
CC differentially expressed gene product in a test sample from a cell
CC suspected of being cancerous, where the gene product is encoded by one
CC of the 5248 polynucleotide sequences given in AAZ12532 to AAZ1779. The
CC polynucleotides can be used as a source of primers and probes, which can
CC be used for a variety of purpose, e.g. detection of expression levels,
CC mapping, tissue typing or profiling, forensics, genetic analysis and
CC detection of polymorphisms. Polypeptides encoded by the polynucleotides
CC can be used for raising antibodies for experimental, diagnostic and
CC therapeutic purposes. The polynucleotides may also be used to construct
CC arrays for diagnostics (which may be used to determine function of an
CC encoded protein); and to detect differences in expression levels between
CC two cells (e.g. to identify abnormal or diseased tissue in a human, to
CC identify a genetic predisposition or susceptibility to a disease such as
CC cancer). The polynucleotides of the invention are especially used in the
CC diagnosis, prognosis and management of colorectal cancer, breast cancer,
CC and lung cancer. The polynucleotides can also be used to screen for
CC peptide analogues and antagonists.
XX
XX Sequence 750 BP; 176 A; 165 C; 156 G; 238 T; 15 other;
SQ

Query Match 52.0%; Score 13; DB 20; Length 750;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gcatgccagcat 13
Db 248 GCATGCCAGCAT 236
|||||

RESULT 31
AAC35170
ID AAC35170 standard; DNA; 974 BP.
XX
XX AAC35170;
AC
XX
XX 17-OCT-2000 (first entry)
DT
XX


```

PR 30-AUG-1999; 99US-01513303.
PR 31-AUG-1999; 99US-01514338.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 52.0%; Score 13; DB 21; Length 974;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 13 ttctgactctcgt 25
Db 593 ttctgactctcgt 605
|||||

RESULT 32
AAC48947
ID AAC48947 standard; DNA; 974 BP.
XX
XX AAC48947;
AC
XX
XX 18-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 59360.
XX
XX Hybridisation assay; genetic mapping; gene expression control;
XX protein identification; signal transduction pathway;
XX metabolic pathway; promoter; termination sequence; ss.
XX
XX Arabidopsis thaliana.
OS

```

```

XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.

```

Wed Oct 10 07:46:03 2001

```
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
```

```
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 13-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 14-OCT-1999; 99US-0159584.
PR 18-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
```

Query Match 52.0%; Score 13; DB 21; Length 974;
Best Local Similarity 100.0%; Pred. No. 1.le-02;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 ttctgatctctcgt 25

|||||

Db 593 ttctgatctcgt 605

RESULT 33

AAC46920/c

ID AAC46920 standard; DNA; 985 BP.

XX

XX AAC46920;

XX

DT 18-OCT-2000 (first entry)

XX

XX Arabidopsis thaliana DNA fragment SEQ ID NO: 51892.

XX

XX Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.

XX

OS Arabidopsis thaliana.

XX

XX EP1033405-A2.

XX

XX 06-SEP-2000.

XX

XX 25-FEB-2000; 2000EP-0301439.

XX

XX

PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123380.
PR 09-MAR-1999; 99US-0123348.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139751.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 30-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0148684.
PR 17-AUG-1999; 99US-0149368.
PR 18-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.


```

PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 52.08; Score 13; DB 21; Length 985;
Best Local Similarity 100.0%; Pred. No. 1.le+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Caps 0;

QY 8 cagcatttcgac 20
    |||||
Db 938 CAGCATTTCGATC 926

RESULT 34
AAC47663/c
ID AAC47663 standard; DNA; 996 BP.
XX
AC AAC47663;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 54655.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 09-MAR-1999; 99US-0123180.
PR 23-MAR-1999; 99US-0123548.
PR 25-MAR-1999; 99US-0125788.
PR 29-MAR-1999; 99US-0126264.
PR 01-APR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.

```

PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 21-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145313.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145919.
PR 02-AUG-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.

PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 52.0%; Score 13; DB 21; Length 996;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 ttctgattccgt 25
|||||
Db 714 TTTCGATCCGCT 702

RESULT 35

AAZ96443
ID AAZ96443 standard; DNA; 1007 BP.

XX AC AAZ96443;

XX DT 10-APR-2000 (first entry)

XX DE S. pneumoniae derived DNA from ORF #271.

XX KW Treatment; prevention; disease; diagnosis; gene therapy; screening;
XX KW bacterial; antimicrobial; antibiotic; pathogenesis; infection; ss.

XX OS Streptococcus pneumoniae.

XX PN WO9806734-A1.

XX PD 19-FEB-1998.

XX PF 15-AUG-1997; 97WO-US14436.

XX PR 16-AUG-1996; 96US-0024022.

XX (SMIK) SMITHKLINE BEECHAM CORP.

XX PI Black Mt, Hodgson JE, Knowles DJC, Lonetto MA, Nicholas RO;
PI Stodola RK;

XX WPI; 1998-159452/14.
XX P-PSDB; AAY86128.

XX PT Streptococcus pneumoniae proteins and related DNA - useful for
XX PT screening compounds for antibacterial activity

XX PS Claim 4; Page 287; 640pp; English.

CC This invention describes novel isolated streptococcus pneumoniae
 CC polynucleotides (see AA296173-296494) and their encoded proteins (see
 CC AA85792-296182). The DNA, vectors and host cells described in the
 CC method of the invention are useful for the recombinant expression of the
 CC polypeptides. The polypeptides are useful for treatment or prevention of
 CC disease, or diagnosis of disease related to expression or activity of
 CC such a polypeptide. They can also be used to screen for compounds which
 CC interact with and inhibit or activate such a polypeptide. The
 CC polypeptides (or DNA encoding them, via gene therapy) are also useful
 CC for inducing an immunological response in a mammal. The antagonists are
 CC useful to inhibit such bacterial polypeptides. The polypeptides are
 CC particularly useful to identify antimicrobial compounds and antibiotics.
 CC They are also useful to determine their role in pathogenesis of
 CC infection, dysfunction and disease.

XX
 SQ Sequence 1007 BP; 327 A; 206 C; 161 G; 313 T; 0 other;

Query Match 52.0%; Score 13; DB 19; Length 1007;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcatgccagcat 13
 |||||
 Db 506 gcatgccagcat 518

RESULT 36
 AAF12052
 ID AAF12052 standard; cDNA; 1038 BP.
 XX
 AC AAF12052;
 XX
 DT 13-MAR-2001 (first entry)
 XX
 DE Aspergillus oryzae EST SEQ ID NO:4575.
 XX
 DE Multiple gene expression; filamentous fungal cell; EST;
 KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
 KW Aspergillus oryzae; trichoderma reesei; identification; recombination;
 KW culture condition; environmental stress; spore morphogenesis;
 KW metabolic pathway engineering; catabolic pathway engineering; ss.
 XX
 OS Aspergillus oryzae.
 XX
 XN WC200056762-A2.
 XX
 PD 28-SEP-2000.
 XX
 XX 22-MAR-2000; 2000WO-US07781.
 XX
 XX 22-MAR-1999; 99US-0273623.
 XX
 XX (NOVO) NOVO NORDISK BIOTECH INC.
 XX (NOVO) NOVO NORDISK AS.
 XX
 XX Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
 XX WPI; 2000-594572/56.
 XX

XX Monitoring differential expression of genes in filamentous fungal cells
 PT uses fluorescence-labeled nucleic acids isolated from the cells and a
 PT substrate of expressed sequence tags -
 XX
 PS Claim 88; Page 1948; 3161pp; English.
 XX
 CC The present invention describes a method for monitoring differential
 CC expression of genes in a first filamentous fungal (FF) cell relative to
 CC expression of the same genes in one or more second filamentous fungal
 CC cells. The method uses fluorescence-labeled nucleic acids isolated from
 CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
 CC are used in the methods for monitoring differential expression of genes
 CC in a first filamentous fungal (FF) cell relative to expression of the

CC same genes in one or more second filamentous fungal cells. Monitoring
 CC the global expression of genes from FF cells allows the production
 CC potential of the microorganisms to be improved. New genes may be
 CC discovered, possible functions of unknown open reading frames can be
 CC identified and gene copy number variation and stability can be
 CC monitored. The expression of genes can be used to study how FF cells
 CC adapt to changes in culture conditions, environmental stress, spore
 CC morphogenesis, recombination, metabolic or catabolic pathway
 CC engineering. Using ESTs provides several advantages over genomic or
 CC random cDNA clones including elimination of redundancy as one spot on an
 CC array equals one gene or open reading frame, and organisation of the
 CC microarrays based on function of the gene products to facilitate
 CC analysis of the results. AAF07478 to AAF11247 represents ESTs from
 CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
 CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
 CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
 CC all specifically claimed in the present invention.

SQ Sequence 1038 BP; 258 A; 294 C; 243 G; 243 T; 0 other;

Query Match 52.0%; Score 13; DB 21; Length 1038;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcatgccagcat 13
 |||||
 Db 593 gcatgccagcat 605

RESULT 37
 AAX84999
 ID AAX84999 standard; DNA; 1065 BP.
 XX
 AC AAX84999;
 XX
 DT 30-JUL-1999 (first entry)
 XX
 DE Human secreted protein gene No. 67.

XX
 DE Human; secreted protein; fusion protein; gene therapy; protein therapy;
 KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
 KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
 KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
 KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
 KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
 KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
 KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.

OS Homo sapiens.
 XX
 XX WO9924836-A1.
 XX
 XX 20-MAY-1999.
 XX
 XX 04-NOV-1998; 98WO-US23435.
 XX
 XX 17-NOV-1997; 97US-0066100.
 PR 07-NOV-1997; 97US-0064900.
 PR 07-NOV-1997; 97US-0064908.
 PR 07-NOV-1997; 97US-0064911.
 PR 07-NOV-1997; 97US-0064912.
 PR 07-NOV-1997; 97US-0064983.
 PR 07-NOV-1997; 97US-0064984.
 PR 07-NOV-1997; 97US-0064985.
 PR 07-NOV-1997; 97US-0064987.
 PR 07-NOV-1997; 97US-0064988.
 PR 17-NOV-1997; 97US-0066090.
 PR 17-NOV-1997; 97US-0066094.
 PR 17-NOV-1997; 97US-0066095.
 PR 17-NOV-1997; 97US-0066089.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.

XX Carter KC, Ebner R, Endress GA, Peng P, Janat F;
 PI Kyaw H, Lafleur DW, Moore PA, Ni J, Olsen HS, Rosen CA;
 PI Ruben SM, Shi Y, Soppet DR, Wei Y;
 XX WPI: 1999-337740/28.
 DR P-PSDB; AAY27633.
 XX
 PT New human secreted proteins and coding sequences useful for treating
 PT disorders of the immune system and hyperproliferative disorders
 XX
 PS Claim 1; Page 316; 507pp; English.
 XX
 CC This sequence represents a nucleic acid molecule which encodes a
 CC secreted human protein. The gene number is given in the descriptor line.
 CC The gene can be used to generate fusion proteins by linking to the gene
 CC to a human immunoglobulin Fc portion (e.g. AAX84924) for increasing the
 CC stability of the fused protein as compared to the human protein only.
 CC The invention relates to 125 novel genes and their fragments (nucleic
 CC acid sequences: AAX84933-X85057; amino acid sequences AAY27567-Y27933)
 CC which are useful for preventing, treating or ameliorating medical
 CC conditions e.g. by protein or gene therapy. Also, pathological
 CC conditions can be diagnosed by determining the amount of the new
 CC polypeptides in a sample or by determining the presence of mutations in
 CC the new polynucleotides. Specific uses are described for each of the 125
 CC polynucleotides, based on which tissues they are most highly expressed in
 CC (see AAX84933 for described uses).
 XX
 SQ Sequence 1065 BP; 217 A; 297 C; 248 G; 303 T; 0 other;

Query Match 52.0%; Score 13; DB 20; Length 1065;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcacgccagcat 13
 |||||
 Db 829 gcacgccagcat 841

RESULT 38
 AAT84197
 ID AAT84197 standard; DNA; 1110 BP.
 XX
 AC AAT84197;
 XX

14-SEP-1998 (first entry)

DNA encoding a UDP-N-acetyl-D-mannosaminuronic acid dehydrogenase.

XX Staphylococcus aureus protein; ribozyme; antisense sequence; control;
 KW Staphylococcal gene; regulatory element; bacterial gene expression;
 KW vaccine; Staphylococcal infection; food poisoning; scaled skin syndrome;
 KW toxic shock syndrome; ss.
 XX

OS Staphylococcus aureus.

XX
 FH Key Location/Qualifiers
 CDS 570..974
 FT /*tag= a
 FT

XX WO9730070-A1.

XX 21-AUG-1997.

XX 19-FEB-1997; 97WO-US02318.

XX 20-FEB-1996; 96US-0011888.

XX (SMIK) SMITHKLINE BEECHAM CORP.

XX Black MT, Burnham MK, Hodgson JE, Knowles DJC, Nicholas RO;

PI Pratt JM, Reichard RW, Rosenberg M, Ward JM;

XX WPI: 1997-424969/39.
 DR P-PSDB; AAW28300.
 XX

PT Novel polypeptide(s) from Staphylococcus aureus strain WCUH29 - used
 PT to isolate antimicrobial compounds, and in vaccines against S.
 PT aureus infection
 XX

PS Claim 9; Page 953; 989pp; English.

XX
 CC The present sequence encodes a Staphylococcus aureus protein, that,
 CC based on homology with an E. coli protein, is believed to be a
 CC UDP-N-acetyl-D-mannosaminuronic acid dehydrogenase (EC 1.1.1.-) (also
 CC known as UDP-mannac dehydrogenase). The present sequence was obtained
 CC from a library of clones of S. aureus WCUH 29 in E. coli. The DNA
 CC sequence can be used in the construction of ribozymes and antisense
 CC sequences to control the expression of Staphylococcal genes. The DNA
 CC sequence is also useful as a source of regulatory elements for the
 CC control of bacterial gene expression. The encoded protein may be used
 CC to produce vaccines to enable a host to produce specific antibodies
 CC with antibacterial action. These vaccines and antibodies would protect
 CC a host against invasion by S. aureus, and conditions relating to
 CC Staphylococcal infection, e.g. Staphylococcal food poisoning, scaled
 CC skin syndrome, and toxic shock syndrome.
 XX

SQ Sequence 1110 BP; 385 A; 152 C; 222 G; 334 T; 17 other;

Query Match 52.0%; Score 13; DB 18; Length 1110;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 atcgccagcattt 15
 |||||
 Db 680 atcgccagcattt 692

RESULT 39
 AAC41953/C
 ID AAC41953 standard; DNA; 1114 BP.
 XX
 AC AAC41953;
 XX

17-OCT-2000 (first entry)

Arabidopsis thaliana DNA fragment SEQ ID NO: 33748.

XX Hybridisation assay; genetic mapping; gene expression control;
 KW protein identification; signal transduction pathway;
 KW metabolic pathway; promoter; termination sequence; ss.

OS Arabidopsis thaliana.

XX EP1033405-A2.

PN 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 23-APR-1999; 99US-0130891.

PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 22-JUL-1999; 99US-0145195.
PR 23-JUL-1999; 99US-0145216.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149802.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 05-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.

PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 52.0%; Score 13; DB 21; Length 1114;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 13 ttctgactctctg 25

Db 674 TTTCGATCCTCGT 662

RESULT 40

AAAL15961

ID AAAL15961 standard; cDNA; 1179 BP.

XX AC AAAL15961;

DT 12-JUN-2000 (first entry)

DE Human protein clone HP10530 coding sequence.

XX Human protein; hydrophobic domain; nutritional source; haematopoiesis;
 KW cytokine production; cell proliferation; cell differentiation;
 KW immune deficiency; infectious disease; autoimmune disorder; asthma;
 KW multiple sclerosis; systemic lupus erythematosus; rheumatoid arthritis;
 KW allergic reaction; osteoporosis; osteoarthritis; periodontal disease;
 KW nervous system disorder; Alzheimer's disease; Parkinson's disease;
 KW Huntington's disease; liver fibrosis; lung fibrosis; reperfusion injury;
 KW systemic cytokine damage; tissue differentiation; contraceptive; stroke;
 KW coagulation disorder; myocardial infarction; inflammatory condition;
 KW septic shock; sepsis; ischaemia; reperfusion injury; arthritis; tumour;
 KW nephritis; therapy; ss.

XX Homo sapiens.

OS WO200005367-A2.

FN 03-FEB-2000.

PD 22-JUL-1999; 99WO-JP03929.

PF 24-JUL-1998; 98JP-0208820.

XX 07-AUG-1998; 98JP-0224105.

PR 25-AUG-1998; 98JP-0238116.

PR 09-SEP-1998; 98JP-0254736.

PR 29-SEP-1998; 98JP-0275505.

XX (SAGA) SAGAMI CHEM RES CENT.
 PA (PROT-) PROTEGENE INC.

XX Kato S, Kimura T;

PI WPI; 2000-182694/16.

DR P-PSDB; AAY9483.

DR

XX Novel human proteins having hydrophobic domains useful for treating
 PT osteoporosis, Alzheimer's disease, Parkinson's disease, asthma,
 PT multiple sclerosis, rheumatoid arthritis, cancer, anaemia, and stroke -
 XX Claim 3; Page 289-290; 351pp; English.

XX This sequence encodes a human protein of the invention, which has
 CC hydrophobic domains. The DNA sequences can be used as a probe or as a
 CC genetic marker. The protein can also be used as a marker, and to identify
 CC potential genetic disorders. The DNA and protein can also be used as
 CC nutritional sources or supplements. The protein exhibits cytokine, cell
 CC proliferation, cell differentiation activities and induces production of
 CC other cytokines in certain cell populations. The protein also exhibits
 CC immune stimulating or immune suppressing activity. It can be used in the
 CC treatment of various immune deficiencies and disorders, and to treat
 CC infectious diseases caused by viral, bacterial, fungal or other
 CC infections. The protein is also used for treating autoimmune disorders
 CC such as multiple sclerosis, systemic lupus erythematosus, and rheumatoid
 CC arthritis. It is also useful in the treatment of allergic reactions and
 CC conditions such as asthma, and in immune suppression after organ
 CC transplantation. The protein is useful in regulation of haematopoiesis
 CC and consequently in the treatment of myeloid or lymphoid cell
 CC deficiencies. It is also used in compositions for tissue growth or
 CC regeneration. The protein is also used in the treatment of osteoporosis
 CC or osteoarthritis and in the treatment of periodontal disease and other
 CC tooth repair processes. The protein is used in the treatment of nervous
 CC system disorders such as Alzheimer's disease, Parkinson's disease, and
 CC Huntington's disease. They are useful for protection or regeneration and
 CC treatment of lung or liver fibrosis, reperfusion injury in various
 CC tissues, and conditions resulting from systemic cytokine damage. They are
 CC also used for promoting or inhibiting tissue differentiation. They are
 CC also used as contraceptives since they exhibit activin or inhibin related
 CC activities and as a fertility inducing therapeutic. They are used for
 CC treating various coagulation disorders and in treatment and prevention of
 CC conditions resulting from coagulation activities e.g. myocardial
 CC infarction or stroke. They also acts as receptors, receptor ligands or
 CC inhibitors or agonists of receptor/ligand interactions. They are used to
 CC treat inflammatory conditions such as septic shock, sepsis, ischaemia
 CC reperfusion injury, arthritis, and nephritis. They can be used to
 CC prevent tumours.

XX Sequence 1179 BP; 244 A; 343 C; 356 G; 236 T; 0 other;

Query Match 52.0%; Score 13; DB 21; Length 1179;

Best Local Similarity 100.0%; Pred. No. 1.1e+02;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ccagcattctcgat 19

Db 549 ccagcattctcgat 561

RESULT 41

AAAC36239/c

ID AAC36239 standard; DNA; 1226 BP.

XX AC AAC36239;

XX 17-OCT-2000 (first entry)

XX Arabidopsis thaliana DNA fragment SEQ ID NO: 13057.

DE Hybridisation assay; genetic mapping; gene expression control;
 KW protein identification; signal transduction pathway;
 KW metabolic pathway; promoter; termination sequence; ss.
 XX Arabidopsis thaliana.

OS EP1033405-A2.

XX EP1033405-A2.

XX 06-SEP-2000.

PD

XX 25-FEB-1999; 99US-0121825.
PF 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 28-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.

PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.

```

PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 52.0%; Score 13; DB 21; Length 1226;
Best Local Similarity 100.0%; Pred. No. 1.le+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 ccagcatttcgat 19
Db 591 CCAGCATTTCGAT 579

RESULT 42
ID AAN82205 standard; DNA; 1263 BP.
AC AAN82205;
XX
XX
DT 12-NOV-1990 (first entry)
DE
DE Nad B gene encoding L aspartate 3' end.
XX
XX Quinolinic acid synthase; L aspartate oxidase; nad A; nad B;
KW plant protection agent; ss.
XX
OS Escherichia coli (K12C600).
XX
XX DE3703255-A.
XX
XX 18-AUG-1988.
XX
XX 04-FEB-1987; 87DE-3703255.
XX
XX 04-FEB-1987; 87DE-3703255.
XX
XX (RUTG ) RURGERSWERKE AG.
XX
XX Laufer A, Gasßen HG, Flachmann R, Hoke H, Holtmann W, Kunz N;
PI

Stadelhofer J, Seifert J;
XX
XX WPI; 1988-235937/34.
XX
XX Plasmids coded for quinolinic acid synthesis - have DNA sequences contg.
PT genetic information for synthesis of quinolinic synthase and L aspartate
PT oxidase.
XX
XX Claim 7; Page 4-5; 24pp; German.
XX
XX Between this 3' end sequence and the 5' end sequence of AAN80036 are
CC ca. 200 bp not represented in the specification.
CC The nad A sequence (AAN80035) encoding a polypeptide with the biological
CC activity of quinolinic acid synthase, together with nad B encoding a
CC polypeptide with the biological activity of L aspartate is inserted into
CC a plasmid. The quinolinic acid expressed is an intermediate for
CC pharmaceuticals and plant protection agents.
XX
XX Sequence 1263 BP; 307 A; 324 C; 344 G; 284 T; 4 other;
SQ

Query Match 52.0%; Score 13; DB 9; Length 1263;
Best Local Similarity 100.0%; Pred. No. 1.le+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 tcgccagcatttc 16
Db 658 tcgccagcatttc 670

RESULT 43
ID AAA93673 standard; DNA; 1287 BP.
AC AAA93673;
XX
XX 16-JAN-2001 (first entry)
DE
DE Human peroxisome associated protein splice variant SECX 3884846-1 DNA.
XX
XX SECX protein; human; secreted; membrane-associated; cancer;
KW proliferation regulator; differentiation regulator; non-malignant tumour;
KW immune disorder; autoimmune disease; transplant rejection; allergy; AIDS;
KW infection; inflammatory disorder; arthritis; haematopoietic disorder;
KW skin disorder; cardiovascular disorder; atherosclerosis; restenosis;
KW neurological disease; Alzheimer's disease; trauma; wound;
KW spinal cord injury; skeletal disorder; cytostatic; immunosuppressive;
KW anti-HIV; antiinflammatory; antiarthritic; antiarteriosclerotic;
KW neuroprotective; vulnery; antiallergic; antimicrobial; cardiant;
KW dermatological; gene therapy; ds.
XX
XX Homo sapiens.
OS
XX
XX WO200053742-A2.
XX
XX 14-SEP-2000.
XX
XX 09-MAR-2000; 2000WO-US06280.
XX
XX 09-MAR-1999; 99US-0123667.
XX 08-MAR-2000; 2000US-0123667.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Shinkets RA;
PI
XX
XX WPI; 2000-594318/56.
XX P-PSDB; AAB23045.
XX
XX Novel human membrane associated or secreted polypeptides and
PT polynucleotides useful for diagnosis, prevention and treatment of
PT pathological states such as cancer, immune, cardiovascular and
PT neurological disorders -

```



```

XX PS Claim 3; Fig 7B; 151pp; English.
XX PS Sequences AAA93616-A93631 and AAA93673-A93676 represent nucleic acids
XX CC which encode human SECX proteins (AAB23029-B23048). The SECX proteins
XX CC of the invention are either secreted or membrane-associated proteins
XX CC and act as regulator of cellular proliferation and differentiation. SECX
XX CC proteins or nucleotides are useful for diagnosing the presence of, or
XX CC predisposition to, a disease associated with altered levels of SECX
XX CC proteins and nucleotides. The SECX proteins are also useful to screen
XX CC compounds that modulate SECX activity or expression. The interaction of
XX CC a SECX protein with other cellular proteins may be useful to modulate
XX CC the activity of a partner protein, cellular proliferation, cellular
XX CC differentiation and cell survival. SECX nucleotides are useful for the
XX CC recombinant expression of SECX protein, and may also be used to detect SECX mRNA
XX CC acid sequences are also useful for identifying a cell or tissue type in
XX CC a biological sample, and in forensic biology. SECX primers or probes are
XX CC useful for detecting the presence of SECX nucleotides and for screening
XX CC tissue cultures for contamination. Diseases that may be treated or
XX CC prevented using SECX proteins or nucleotides include cancer (e.g.,
XX CC colorectal carcinoma, prostate cancer), benign tumours, immune disorders
XX CC (including autoimmune diseases, transplant rejection, allergies, AIDS),
XX CC infections, inflammatory disorders, arthritis, haematopoietic disorders,
XX CC skin disorders, cardiovascular disorders, atherosclerosis, restenosis,
XX CC neurological diseases (e.g., Alzheimer's disease), trauma (e.g.,
XX CC surgical or traumatic wounds, spinal cord injury), and skeletal
XX CC disorders.
XX SQ Sequence 1287 BP; 266 A; 368 C; 388 G; 262 T; 3 other;

Query Match 52.0%; Score 13; DB 21; Length 1287;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 ccagcattcgat 19
Db 609 ccagcattcgat 621
|||||

RESULT 44
AAA93674
ID AAA93674 standard; DNA; 1291 BP.
XX AC AAA93674;
XX DT 16-JAN-2001 (first entry)
XX DE Human peroxisome associated protein splice variant SECX 3884846-2 DNA.
XX KW SECX protein; human; secreted; membrane-associated; cancer;
XX KW proliferation regulator; differentiation regulator; non-malignant tumour;
XX KW immune disorder; autoimmune disease; transplant rejection; allergy; AIDS;
XX KW infection; inflammatory disorder; arthritis; haematopoietic disorder;
XX KW skin disorder; cardiovascular disorder; atherosclerosis; restenosis;
XX KW neurological disease; Alzheimer's disease; trauma; woundings;
XX KW spinal cord injury; skeletal disorder; cytostatic; immunosuppressive;
XX KW anti-HIV; antiinflammatory; antiarthritic; antiarteriosclerotic;
XX KW neuroprotective; vulnary; antiallergic; antimicrobial; cardiant;
XX KW dermatological; gene therapy; ds.
XX OS Homo sapiens.
XX PN WO200053742-A2.
XX PD 14-SEP-2000.
XX PF 09-MAR-2000; 2000WO-US06280.
XX PR 09-MAR-1999; 99US-0123667.
XX PR 08-MAR-2000; 2000US-0123667.

XX PA (CURA-) CURAGEN CORP.
XX PI Shimkets RA;
XX DR WPI: 2000-594318/56.
XX DR P-PSDB; AAB23046.
XX PT Novel human membrane associated or secreted polypeptides and
XX PT polynucleotides useful for diagnosis, prevention and treatment of
XX PT pathological states such as cancer, immune, cardiovascular and
XX PT neurological disorders
XX PS Claim 3; Fig 7C; 151pp; English.
XX PS Sequences AAA93616-A93631 and AAA93673-A93676 represent nucleic acids
XX CC which encode human SECX proteins (AAB23029-B23048). The SECX proteins
XX CC of the invention are either secreted or membrane-associated proteins
XX CC and act as regulator of cellular proliferation and differentiation. SECX
XX CC proteins or nucleotides are useful for diagnosing the presence of, or
XX CC predisposition to, a disease associated with altered levels of SECX
XX CC proteins and nucleotides. The SECX proteins are also useful to screen
XX CC compounds that modulate SECX activity or expression. The interaction of
XX CC a SECX protein with other cellular proteins may be useful to modulate
XX CC the activity of a partner protein, cellular proliferation, cellular
XX CC differentiation and cell survival. SECX nucleotides are useful for the
XX CC recombinant expression of SECX protein, and may also be used to detect SECX mRNA
XX CC acid sequences are also useful for identifying a cell or tissue type in
XX CC a biological sample, and in forensic biology. SECX primers or probes are
XX CC useful for detecting the presence of SECX nucleotides and for screening
XX CC tissue cultures for contamination. Diseases that may be treated or
XX CC prevented using SECX proteins or nucleotides include cancer (e.g.,
XX CC colorectal carcinoma, prostate cancer), benign tumours, immune disorders
XX CC (including autoimmune diseases, transplant rejection, allergies, AIDS),
XX CC infections, inflammatory disorders, arthritis, haematopoietic disorders,
XX CC skin disorders, cardiovascular disorders, atherosclerosis, restenosis,
XX CC neurological diseases (e.g., Alzheimer's disease), trauma (e.g.,
XX CC surgical or traumatic wounds, spinal cord injury), and skeletal
XX CC disorders.
XX SQ Sequence 1291 BP; 267 A; 370 C; 389 G; 262 T; 3 other;

Query Match 52.0%; Score 13; DB 21; Length 1291;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 ccagcattcgat 19
Db 613 ccagcattcgat 625
|||||

RESULT 45
AAC61879
ID AAC61879 standard; cDNA; 1313 BP.
XX AC AAC61879;
XX AC AAC61879;
XX XX 06-MAR-2001 (first entry)
XX DT
XX DE cDNA encoding a human secreted protein.
XX OS
XX KW Secreted protein; platelet disorder; stem cell disorder; osteoporosis;
XX KW osteoarthritis; burn; incision; ulcer; periodontal disease; neuropathy;
XX KW nervous system disease; bone growth; cosmetic plastic surgery;
XX KW gut protection; gut regeneration; fibrosis; cancer;
XX KW bone marrow transplantation; ss.
XX XX Homo sapiens.
XX OS
XX FH Key Location/Qualifiers

```

```

FT CDS      163..1140
FT          /*tag= a
FT          /product= "secreted protein"
XX PN      WO200061755-A2.
XX PD      19-OCT-2000.
XX PF      10-APR-2000; 2000WO-US09555.
XX PR      09-APR-1999; 99US-0128574.
XX PR      20-AUG-1999; 99US-0150054.
XX PA      (CHIR ) CHIRON CORP.
XX PI      Garcia PD;
XX DR      WPI; 2000-665133/64.
XX DR      P-PSDB; AAB19393.
XX PT      Novel secreted human proteins useful for stimulating blood cell
XX PT      generation in patients receiving cancer chemotherapy, treating bone
XX PT      marrow transplantation patients and for healing fractured bones
XX PS      Claim 2; Page 54; 74pp; English.
XX CC      AAC61879-93 encode secreted human proteins. The secreted proteins are
XX CC      useful in assays to determine their biological activities. The proteins
XX CC      can also be used as biomarkers to identify tissues or cell types which
XX CC      express the proteins. The polynucleotide molecules can be used as
XX CC      biomarkers for tissues or chromosomes and to elicit immune responses.
XX CC      The proteins and antibodies are useful in diagnosis and treatment of
XX CC      diseases associated with altered expression of these proteins. The
XX CC      proteins are also useful for prevention or treatment of platelet
XX CC      disorders, stem cell disorders, osteoporosis or osteoarthritis, burns,
XX CC      incisions, ulcers, periodontal diseases, central and peripheral nervous
XX CC      system diseases and neuropathies, for healing fractured bones and to
XX CC      induce cartilage and/or bone growth in cosmetic plastic surgery. The
XX CC      proteins are also useful for gut protection or regeneration. The
XX CC      treatment of lung or liver fibrosis, for stimulating blood cell
XX CC      generation in patients receiving cancer chemotherapy and for treatment
XX CC      of bone marrow transplantation patients.
XX SQ      Sequence 1313 BP; 298 A; 366 C; 360 G; 289 T; 0 other;

Query Match      52.0%; Score 13; DB 21; Length 1313;
Best Local Similarity 100.0%; Pred. No. 1.le+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      7 ccagcatttcgat 19
Db      711 ccagcatttcgat 723

```

Search completed: October 9, 2001, 15:52:46
 Job time: 15127 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 15:55:36 ; Search time 218.82 Seconds
(without alignments)
21.629 Million cell updates/sec

Title: US-09-396-196F-7

Perfect score: 25

Sequence: 1 gcatcgccagcattgcctcgt 25

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 324599 seqs, 94655562 residues

Word size : 9

Total number of hits satisfying chosen parameters: 8423

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents.NA.*

- 1: /cgn2_6/ptodata/2/ina/5A.COMB.seq.*
- 2: /cgn2_6/ptodata/2/ina/5B.COMB.seq.*
- 3: /cgn2_6/ptodata/2/ina/6A.COMB.seq.*
- 4: /cgn2_6/ptodata/2/ina/6B.COMB.seq.*
- 5: /cgn2_6/ptodata/2/ina/PCTUS.COMB.seq.*
- 6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	25	100.0	1041	2	US-08-401-068-7
2	25	100.0	1041	2	US-08-846-338-7
3	25	100.0	5872	3	US-08-411-768B-1
4	25	100.0	5872	3	US-08-411-768B-6
5	14	56.0	2232	5	PCT-US96-05320A-113
6	14	56.0	5319	1	US-08-169-927-1
7	13	52.0	61	4	US-08-836-075A-51
8	13	52.0	447	4	US-08-933-750C-60
9	13	52.0	1460	2	US-09-234-613-60
10	13	52.0	1460	2	US-08-743-637B-172
11	13	52.0	1650	3	US-08-526-840B-172
12	13	52.0	2104	1	US-08-682-193A-1
13	13	52.0	2193	1	US-08-420-235B-16
14	13	52.0	2193	1	US-08-793-624-16
15	13	52.0	2193	5	PCT-US95-10194-16
16	13	52.0	2312	2	US-09-134-566-15
17	13	52.0	2312	2	US-09-134-566-15
18	13	52.0	2312	2	US-09-134-566-15
19	13	52.0	2973	2	US-08-343-101A-5
20	13	52.0	2973	3	US-09-183-688-5
21	13	52.0	20710	1	US-08-420-235B-1
22	13	52.0	20710	3	US-08-793-624-1
23	13	52.0	20710	5	PCT-US95-10194-1
24	13	52.0	35100	2	US-08-770-379-18
25	13	52.0	35100	4	US-08-757-669A-18
26	12	48.0	19	3	US-08-851-843A-92
27	12	48.0	19	4	US-08-974-549A-384

c 28 12 48.0 38 1 US-08-373-124A-1692 Sequence 1692, Ap
c 29 12 48.0 38 1 US-08-435-628-1692 Sequence 1692, Ap
c 30 12 48.0 660 2 US-08-743-637B-175 Sequence 175, App
c 31 12 48.0 660 3 US-08-526-840B-175 Sequence 175, App
32 12 48.0 715 4 US-08-998-416-767 Sequence 767, App
33 12 48.0 780 3 US-08-927-128-3 Sequence 3, Appli
34 12 48.0 841 4 US-08-990-823-39 Sequence 39, Appli
35 12 48.0 864 2 US-08-595-868C-11 Sequence 11, Appli
36 12 48.0 867 4 US-09-139-819A-11 Sequence 11, Appli
c 37 12 48.0 954 1 US-08-624-125-9 Sequence 9, Appli
c 38 12 48.0 1188 1 US-08-781-562-2 Sequence 2, Appli
c 39 12 48.0 1239 4 US-09-064-693A-20 Sequence 20, Appli
40 12 48.0 1249 3 US-08-985-950-1 Sequence 1, Appli
c 41 12 48.0 1305 2 US-08-484-575A-19 Sequence 19, Appli
c 42 12 48.0 1305 3 US-08-477-459-19 Sequence 19, Appli
c 43 12 48.0 1305 3 US-08-479-869-19 Sequence 19, Appli
c 44 12 48.0 1305 3 US-08-486-414-19 Sequence 19, Appli
c 45 12 48.0 1305 5 PCT-US94-01826A-19 Sequence 19, Appli

ALIGNMENTS

RESULT 1
US-08-401-068-7
; Sequence 7, Application US/08401068
; Patent No. 5859335
; GENERAL INFORMATION:
; APPLICANT: Patton, David
; TITLE OF INVENTION: Enhanced Biotin Biosynthesis in Plant Tissue
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA: US/08401,068
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/351,970
; FILING DATE: 08-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1041 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEetical: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1038
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /product= "biotin synthase"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
US-08-401-068-7

Query Match 100.0%; Score 25; DB 2; Length 1041;
 Best Local Similarity 100.0%; Pred. No. 9.3e-06;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcacgccagcatttcgattcctcgt 25
 |||
 DB 90 GCATGCCAGCATTTCGATCCTCGT 114

RESULT 2

US-08-846-338-7
 ; Sequence 7, Application US/08846338
 ; Patent No. 5869719
 ; GENERAL INFORMATION:
 ; APPLICANT: Patton, David
 ; TITLE OF INVENTION: Transgenic Plants Having Increased Biotin Content
 ; NUMBER OF SEQUENCES: 20
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: No. 5869719artis Corporation
 ; STREET: 520 White Plains Road, P.O. Box 2005
 ; CITY: Tarrytown
 ; STATE: NY
 ; COUNTRY: USA
 ; ZIP: 10591
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30B
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/846,338
 ; FILING DATE:
 ; CLASSIFICATION: 800
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Meigs, J. Timothy
 ; REGISTRATION NUMBER: 38,241
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 919-541-8587
 ; TELEFAX: 919-541-8689
 ; INFORMATION FOR SEQ ID NO: 7:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1041 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; HYPOTHETICAL: NO
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 1..1038
 ; IDENTIFICATION METHOD: experimental
 ; OTHER INFORMATION: /product= "biotin synthase"
 ; OTHER INFORMATION: /evidence= EXPERIMENTAL
 ; US-08-846-338-7

Query Match 100.0%; Score 25; DB 2; Length 1041;
 Best Local Similarity 100.0%; Pred. No. 9.3e-06;
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcacgccagcatttcgattcctcgt 25
 |||
 DB 90 GCATGCCAGCATTTCGATCCTCGT 114

RESULT 3

US-08-411-768B-1
 ; Sequence 1, Application US/08411768B
 ; Patent No. 6083712
 ; GENERAL INFORMATION:
 ; APPLICANT: Olwen Birch
 ; APPLICANT: Johann Brass
 ; APPLICANT: Martin Fuhrmann

APPLICANT: Nicholas Shaw
 TITLE OF INVENTION: Biotechnological Method
 OF PRODUCING BIOTIN
 NUMBER OF SEQUENCES: 19
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
 STREET: 30 Rockefeller Plaza
 CITY: New York
 STATE: New York
 COUNTRY: USA
 ZIP: 10112
 COMPUTER READABLE FORM:
 MEDIUM TYPE: floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: WordPerfect
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/411,768B
 FILING DATE: 31-March-95
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: CH 3124/92
 FILING DATE: 02-OCT-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: CH 2134/93
 FILING DATE: 15-JUL-1993
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 5872 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 ORIGINAL SOURCE:
 ORGANISM: Escherichia coli
 STRAIN: DSM498
 IMMEDIATE SOURCE:
 CLONE: pB030A-15/9
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 117..1157
 IDENTIFICATION METHOD: experimental
 OTHER INFORMATION: /codon_start= 117
 OTHER INFORMATION: /product= "biotin synthase"
 OTHER INFORMATION: /evidence= EXPERIMENTAL
 OTHER INFORMATION: /gene= "bioB"
 OTHER INFORMATION: /number= 1
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 2295..3050
 OTHER INFORMATION: /codon_start= 2295
 OTHER INFORMATION: /function= "involved in pimeloyl-CoA synthesis"
 OTHER INFORMATION: /product= "protein"
 OTHER INFORMATION: /gene= "bioc"
 OTHER INFORMATION: /number= 3
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 3750..5039
 IDENTIFICATION METHOD: experimental
 OTHER INFORMATION: /codon_start= 3750
 OTHER INFORMATION: /EC_number= 2.6.1.62
 OTHER INFORMATION: /product= "DAPA synthase"
 OTHER INFORMATION: /evidence= EXPERIMENTAL
 OTHER INFORMATION: /gene= "bioA"
 OTHER INFORMATION: /number= 5
 OTHER INFORMATION: /standard_name= "S-Adenosyl-L-methionine:8-amino-7-oxononanoate
 OTHER INFORMATION: aminotransf."
 FEATURE:
 NAME/KEY: CDS

```

; LOCATION: 5098..5574
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /codon_start= 5098
; OTHER INFORMATION: /function= "unknown, involved in biotin synthesis"
; OTHER INFORMATION: /product= "protein"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /gene= "ORF1"
; OTHER INFORMATION: /number= 6
; FEATURE:
; NAME/KEY: -10_signal
; LOCATION: 45..49
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /standard_name= "promoter ptac"
; FEATURE:
; NAME/KEY: -35_signal
; LOCATION: 23..28
; OTHER INFORMATION: /standard_name= "promoter ptac"
; FEATURE:
; NAME/KEY: RBS
; LOCATION: 105..119
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /standard_name= "bioB RBS no.9"
; FEATURE:
; NAME/KEY: RBS
; LOCATION: 2284..2297
; OTHER INFORMATION: /standard_name= "bioC RBS"
; FEATURE:
; NAME/KEY: RBS
; LOCATION: 3742..3752
; OTHER INFORMATION: /standard_name= "bioA RBS"
; FEATURE:
; NAME/KEY: RBS
; LOCATION: 5088..5100
; OTHER INFORMATION: /standard_name= "ORF1 RBS"
; FEATURE:
; NAME/KEY: terminator
; LOCATION: 5583..5644
; OTHER INFORMATION: /standard_name= "rho-independent
; OTHER INFORMATION: transcriptional terminator"
; FEATURE:
; NAME/KEY: stem_loop
; LOCATION: 5583..5605
; FEATURE:
; NAME/KEY: promoter
; LOCATION: 1..96
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /function= "promoter ptac"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION:
; PUBLICATION INFORMATION:
; DOCUMENT NUMBER: WO 87/01391 B1
; FILING DATE: 26-AUG-1986
; PUBLICATION DATE: 07-APR-1993
; US-08-411-768B-1

Query Match 100.0%; Score 25; DB 3; Length 5872;
Best Local Similarity 100.0%; Pred. No. 8.9e-06;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcatgccagcatttcgactcgt 25
|||||
DB 206 GCATGCCAGCATTTCGATCCTCGT 230

RESULT 4
US-08-411-768B-6
; Sequence 6, Application US/08411768B
; Patent No. 6083712
; GENERAL INFORMATION:
; APPLICANT: Olwen Birch
; APPLICANT: Johann Brass

```

```

; APPLICANT: Martin Fuhrmann
; APPLICANT: Nicholas Shaw
; TITLE OF INVENTION: Biotechnological Method
; TITLE OF INVENTION: of Producing Biotin
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect
; SOFTWARE: Version 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/411,768B
; FILING DATE: 31-March-95
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 3124/92
; FILING DATE: 02-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: CH 2134/93
; FILING DATE: 15-JUL-1993
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5872 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli
; STRAIN: DSM498
; IMMEDIATE SOURCE:
; CLONE: pBO30A15-9
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1154..2308
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /codon_start= 1154
; OTHER INFORMATION: /EC_number= 2.3.1.47
; OTHER INFORMATION: /product= "KAPA synthase"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /gene= "bioF"
; OTHER INFORMATION: /number= 2
; OTHER INFORMATION: /standard_name= "8-Amino-7-oxononanoate synthase"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 3043..3753
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /codon_start= 3043
; OTHER INFORMATION: /EC_number= 6.3.3.3
; OTHER INFORMATION: /product= "DTB synthase"
; OTHER INFORMATION: /evidence= EXPERIMENTAL
; OTHER INFORMATION: /gene= "bioD"
; OTHER INFORMATION: /number= 4
; OTHER INFORMATION: /standard_name= "Dethiobiotin synthase"
; FEATURE:
; NAME/KEY: RBS
; LOCATION: 1141..1156
; OTHER INFORMATION: /standard_name= "bioF RBS"
; FEATURE:
; NAME/KEY: RBS
; LOCATION: 3030..3045
; OTHER INFORMATION: /standard_name= "bioD RBS"
; PUBLICATION INFORMATION:
; DOCUMENT NUMBER: WO 87/01391 B1

```

;
; FILING DATE: 26-AUG-1986
; PUBLICATION DATE: 07-APR-1993
; US-08-411-768B-6

Query Match 100.0%; Score 25; DB 3; Length 5872;
Best Local Similarity 100.0%; Pred. No. 8.9e-06;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 qcatgccagcatttcgactcgt 25
Db 206 GCATGCCAGCATTTCGACTCTCGT 230

RESULT 5

PCT-US96-05320A-113/c
; Sequence 113, Application PC/TUS9605320A
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences
; APPLICANT: 9410 Key West Avenue
; APPLICANT: Rockville, MD 20850
; APPLICANT: United States of America
; APPLICANT: Johns Hopkins University
; APPLICANT: 720 Rutland Avenue
; APPLICANT: Baltimore, MD 21205
; APPLICANT: United States of America
; APPLICANT: Mark D. Adams
; APPLICANT: Owen White
; APPLICANT: Hamilton O. Smith
; APPLICANT: J. Craig Venter
; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus Influenzae Rd Genome,
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20003-3934

COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/05320A
; FILING DATE: April 22, 1996

CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/476,102
; FILING DATE: June 7, 1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/487,429
; FILING DATE: June 7, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Eric K. Steffe
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488.014PC01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540

INFORMATION FOR SEQ ID NO: 113:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2232 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
PCT-US96-05320A-113

Query Match 56.0%; Score 14; DB 5; Length 2232;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 atcgcagcatttc 16
Db 636 ATCGCAGCATTTC 623

RESULT 6

US-08-169-927-1/c
; Sequence 1, Application US/08169927
; Patent No. 5783441
; GENERAL INFORMATION:
; APPLICANT: Carl, Mitchell
; APPLICANT: Dobson, Michael E.
; APPLICANT: Ching, Wei Mei
; APPLICANT: Daschl, Gregory A
; TITLE OF INVENTION: Gene and Protein Applicable to the
; TITLE OF INVENTION: Preparation of Vaccines for Rickettsia prowazekii and
; TITLE OF INVENTION: Rickettsia typhi and the Detection of Both
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Counsel, Naval Medical R & D Command
; STREET: Bldg. 1, T-12, 8901 Wisconsin Ave.
; CITY: Bethesda
; STATE: MD
; COUNTRY: USA
; ZIP: 20889-5606

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/169,927
; FILING DATE:

CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/742,128
; FILING DATE: 08/09/91
; ATTORNEY/AGENT INFORMATION:
; NAME: Spevack, A. David
; REGISTRATION NUMBER: 24,743
; REFERENCE/DOCKET NUMBER: 75,976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 295-6759
; TELEFAX: (301) 295-1022

INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5319 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; ORGANISM: Rickettsia prowazekii
; STRAIN: Breinl
; FEATURE:

NAME/KEY: -35_signal
LOCATION: 340..345
; FEATURE:
NAME/KEY: -10_signal
LOCATION: 363..368
; FEATURE:
NAME/KEY: CDS
LOCATION: 391..5226
; FEATURE:
NAME/KEY: RBS
LOCATION: 379..386
; FEATURE:
NAME/KEY: stem_loop
LOCATION: 5270..5306

COUNTRY: US
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: WordPerfect 8.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/023,228B
 FILING DATE: 12-FEBRUARY-1998
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: PCT/US97/01739
 FILING DATE: 30 JAN 1997
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/595,335
 FILING DATE: 1 FEB 1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Barry J. Swanson, Esq.
 REGISTRATION NUMBER: 33,215
 REFERENCE/DOCKET NUMBER: NEX50/CIP
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (303) 793-3333
 TELEFAX: (303) 793-3433
 INFORMATION FOR SEQ ID NO: 59:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 61 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: RNA
 FEATURE:
 OTHER INFORMATION: All pyrimidines are 2'-F modified
 US-09-023-228B-59

```

; APPLICATION NUMBER: EP 95870076.7
; FILING DATE: 28 Jun 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: KAMMERER, PATRICIA A.
; REGISTRATION NUMBER: 29,775
; REFERENCE/DOCKET NUMBER: INNS:004
; INFORMATION FOR SEQ ID NO: 51:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 447 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-836-075A-51

Query Match 52.0%; Score 13; DB 4; Length 447;
Best Local Similarity 100.0%; Pred. NO. 43;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 catgccagcatt 14
| | | | | | | | | |
Db 406 CATGCCAGCATT 418

RESULT 9
US-08-933-750C-60
Sequence 60, Application US/08933750C

```

```

; Patent No. 5932442
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Bandman, Olga
; APPLICANT: Shah, Purvi
; APPLICANT: Au-Young, Janice
; APPLICANT: Yue, Henry
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
; NUMBER OF SEQUENCES: 98
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/933,750C
; FILING DATE: September 23, 1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0356 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 60:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1460 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: TLYMNOR01
; CLONE: 140516
; US-09-933-750C-60

```

```

Query Match 52.0%; Score 13; DB 2; Length 1460;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ccagcatttcgat 19
Db 722 CCAGCATTTCGAT 734

```

```

RESULT 10
US-09-234-613-60
; Sequence 60, Application US/09234613
; Patent No. 6132973
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Bandman, Olga
; APPLICANT: Shah, Purvi
; APPLICANT: Au-Young, Janice
; APPLICANT: Yue, Henry
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.

```

```

; TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
; NUMBER OF SEQUENCES: 98
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/234,613
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/933,750
; FILING DATE: September 23, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0356 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 60:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1460 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: TLYMNOR01
; CLONE: 140516
; US-09-234-613-60

```

```

Query Match 52.0%; Score 13; DB 3; Length 1460;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ccagcatttcgat 19
Db 722 CCAGCATTTCGAT 734

```

```

RESULT 11
US-08-743-637B-172/c
; Sequence 172, Application US/08743637B
; Patent No. 5994066
; GENERAL INFORMATION:
; APPLICANT: BERGERON, Michel G.
; APPLICANT: PICARD, Francois J.
; APPLICANT: OUELLETTE, Marc
; APPLICANT: ROY, Paul H.
; TITLE OF INVENTION: SPECIES-SPECIFIC AND UNIVERSAL DNA
; TITLE OF INVENTION: PROBES AND AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND
; TITLE OF INVENTION: IDENTIFY COMMON BACTERIAL PATHOGENS AND ASSOCIATED
; TITLE OF INVENTION: ANTIBIOTIC RESISTANCE GENES FROM CLINICAL SPECIMENS ...
; NUMBER OF SEQUENCES: 273
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: QUARLES & BRADY
; STREET: 411 EAST WISCONSIN AVENUE
; CITY: MILWAUKEE
; STATE: WISCONSIN
; COUNTRY: USA
; ZIP: 53202-4497
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```



```
;
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/743,637B
; FILING DATE: 04-NOV-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/526,840
; FILING DATE: 11-SEP-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: BAKER, Jean C.
; REGISTRATION NUMBER: 35,433
; REFERENCE/DOCKET NUMBER: 850586.90012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (414) 277-5000
; TELEFAX: (414) 277-5591
; INFORMATION FOR SEQ ID NO: 172:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1650 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-743-637B-172

Query Match          52.0%; Score 13; DB 2; Length 1650;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 cagcattcgatc 20
   |||||
Db 555 CAGCATTTCGATC 543

RESULT 12
US-08-526-840B-172/c
; Sequence 172, Application US/08526840B
; Patent No. 6001564
; GENERAL INFORMATION:
; APPLICANT: BERGERON, Michel G.
; APPLICANT: QUELLETTE, Marc
; APPLICANT: ROY, Paul H.
; TITLE OF INVENTION: SPECIFIC AND UNIVERSAL PROBES AND
; TITLE OF INVENTION: APPLICATION PRIMERS TO RAPIDLY DETECT AND IDENTIFY
; TITLE OF INVENTION: COMMON BACTERIAL PATHOGENS AND ANTIBIOTIC RESISTANCE GENES
; TITLE OF INVENTION: FROM CLINICAL SPECIMENS FOR ROUTINE DIAGNOSIS IN ...
; NUMBER OF SEQUENCES: 177
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: QUARLES & BRADY
; STREET: 411 East Wisconsin Avenue
; CITY: Milwaukee
; STATE: Wisconsin
; COUNTRY: USA
; ZIP: 53202-4497
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/526,840B
; FILING DATE: 11-SEP-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/304,732
; FILING DATE: 12-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BAKER, Jean C.
; REGISTRATION NUMBER: 35,433
; REFERENCE/DOCKET NUMBER: 850586.90012
; TELECOMMUNICATION INFORMATION:
```

```
;
; TELEPHONE: (414) 277-5000
; TELEFAX: (414) 277-5591
; INFORMATION FOR SEQ ID NO: 172:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1650 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-526-840B-172

Query Match          52.0%; Score 13; DB 3; Length 1650;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 cagcattcgatc 20
   |||||
Db 555 CAGCATTTCGATC 543

RESULT 13
US-08-682-193A-1
; Sequence 1, Application US/08682193A
; Patent No. 5776740
; GENERAL INFORMATION:
; APPLICANT: HATAKEYAMA, Kazuhisa
; APPLICANT: GOTO, Makoto
; APPLICANT: TERASAWA, Masato
; APPLICANT: YUKAWA, Hideaki
; TITLE OF INVENTION: PROCESS FOR THE PREPARATION OF
; TITLE OF INVENTION: L-TRYPTOPHANE
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSHMAN, LLP
; STREET: 130 WATER STREET
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02019
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/682,193A
; FILING DATE: 17-JUL-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 7-181730
; FILING DATE: 18-JUL-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: DAVID, RESNICK S
; REGISTRATION NUMBER: 34,235
; REFERENCE/DOCKET NUMBER: 46643
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-523-3400
; TELEFAX: 617-523-6440
; TELEX: 200291 STRE
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2104 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; ORGANISM: Brevibacterium flavum
; STRAIN: MJ-233
```

;
;
;
; NAME/KEY: Coding Sequence
; LOCATION: 556...1855
; OTHER INFORMATION:
US-08-682-193A-1

Query Match 52.0%; Score 13; DB 1; Length 2104;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 13 ttatgactcttgt 25
|||||
Db 1635 TTTCGATCCTCGT 1647

RESULT 14
US-08-420-235B-16/C
; Sequence 16, Application US/08420235B
; Patent No. 5801042
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
; TITLE OF INVENTION: SEQUENCES AND USES THEREOF
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/420,235B
FILING DATE:
CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 45185-B
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 2193 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: N
ANTI-SENSE: N
FEATURE:
NAME/KEY: CDS
LOCATION: 1..2193
OTHER INFORMATION:
US-08-420-235B-16

Query Match 52.0%; Score 13; DB 1; Length 2193;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gcatgccagcat 13
|||||
Db 45 GCATGCCAGCAT 33

RESULT 15
US-08-793-624-16/c
; Sequence 16, Application US/08793624C
; Patent No. 6150093
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Moore, Patrick S.

; TITLE OF INVENTION: Unique Associated Kaposi's Sarcoma Virus Sequences And
; TITLE OF INVENTION: Uses Thereof
; FILE REFERENCE: 45185-C-PCT-US/JPW
; CURRENT APPLICATION NUMBER: US/08/793,624C
; CURRENT FILING DATE: 1997-02-18
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 16
; LENGTH: 2193
; TYPE: DNA
; ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-08-793-624-16

Query Match 52.0%; Score 13; DB 3; Length 2193;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gcatgccagcat 13
|||||
Db 45 GCATGCCAGCAT 33

RESULT 16
PCT-US95-10194-16/C
; Sequence 16, Application PC/TUS9510194
; GENERAL INFORMATION:
; APPLICANT: The Trustees of Columbia University in the City of New York
; APPLICANT: City
; TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
; TITLE OF INVENTION: SEQUENCES AND USES THEREOF
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/10194
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 45185-C-PCT/JPW/NSC
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 2193 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: N
ANTI-SENSE: N
FEATURE:

; NAME/KEY: CDS
; LOCATION: 1..2193
; OTHER INFORMATION:
PCT-US95-10194-16

Query Match 52.0%; Score 13; DB 5; Length 2193;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcatgccagcat 13
|||||
Db 45 GCATGCCAGCAT 33

RESULT 17

US-09-134-566-15
; Sequence 15, Application US/09134566
; Patent No. 5998147

GENERAL INFORMATION:

; APPLICANT: Petit, Christine
; APPLICANT: Denoyelle-Gryson, Francoise
; APPLICANT: Weil, Dominique
; APPLICANT: Marlin-Duvernois, Sandrine
; APPLICANT: Guesdon, Jean-Luc
; TITLE OF INVENTION: MUTATED POLYNUCLEOTIDE CORRESPONDING TO A MUTATION
; TITLE OF INVENTION: RESPONSIBLE FOR PRELINGUAL NON-SYNDROMIC DEAFNESS IN
; TITLE OF INVENTION: THE CONNEXIN 26 GENE AND METHOD OF DETECTING THIS
; TITLE OF INVENTION: HEREDITARY DEFECT
; FILE REFERENCE: 03495-0175
; CURRENT APPLICATION NUMBER: US/09/134,566
; CURRENT FILING DATE: 1998-08-14
; EARLIER APPLICATION NUMBER: 60/055,863
; EARLIER FILING DATE: 1998-08-15
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
; LENGTH: 2312
; TYPE: DNA
; ORGANISM: human connexin
US-09-134-566-15

Query Match 52.0%; Score 13; DB 2; Length 2312;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 tcgccagcat 16
|||||
Db 960 tcgccagcat 972

RESULT 18

US-09-134-566-8
; Sequence 8, Application US/09134566
; Patent No. 5998147

GENERAL INFORMATION:

; APPLICANT: Petit, Christine
; APPLICANT: Denoyelle-Gryson, Francoise
; APPLICANT: Weil, Dominique
; APPLICANT: Marlin-Duvernois, Sandrine
; APPLICANT: Guesdon, Jean-Luc
; TITLE OF INVENTION: MUTATED POLYNUCLEOTIDE CORRESPONDING TO A MUTATION
; TITLE OF INVENTION: RESPONSIBLE FOR PRELINGUAL NON-SYNDROMIC DEAFNESS IN
; TITLE OF INVENTION: THE CONNEXIN 26 GENE AND METHOD OF DETECTING THIS
; TITLE OF INVENTION: HEREDITARY DEFECT
; FILE REFERENCE: 03495-0175
; CURRENT APPLICATION NUMBER: US/09/134,566
; CURRENT FILING DATE: 1998-08-14
; EARLIER APPLICATION NUMBER: 60/055,863
; EARLIER FILING DATE: 1998-08-15
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 8
; LENGTH: 2314
; TYPE: DNA
; ORGANISM: human connexin
US-09-134-566-8

Query Match 52.0%; Score 13; DB 2; Length 2314;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 tcgccagcat 16
|||||
Db 962 tcgccagcat 974

RESULT 19

US-08-343-101A-5/c
; Sequence 5, Application US/08343101A
; Patent No. 5830759

GENERAL INFORMATION:

; APPLICANT: Chang, Yuan
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: Unique Associated Kaposi's Sarcoma
; TITLE OF INVENTION: Virus Sequences And Uses Thereof
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/343,101A
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: White Esq., John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 45185-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-278-0400
; TELEFAX: 212-391-0526
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2973 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: N
; ANTI-SENSE: N
US-08-343-101A-5

Query Match 52.0%; Score 13; DB 2; Length 2973;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcatgccagcat 13
|||||
Db 2181 GCATGCCAGCAT 2169

RESULT 20

US-09-183-688-5/c
; Sequence 5, Application US/09183688

; Patent No. 6093550
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: Unique Associated Kaposi's Sarcoma
; TITLE OF INVENTION: Virus Sequences And Uses Thereof
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/183,688
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/343,101
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: White Esq., John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 45185-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-278-0400
; TELEFAX: 212-391-0526
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2973 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: N
; ANTI-SENSE: N
US-09-183-688-5

Query Match 52.0%; Score 13; DB 3; Length 2973;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gcatcgccagcat 13
|||||
Db 2181 GCATCGCCAGCAT 2169

RESULT 21
US-08-420-235B-1/c
; Sequence 1, Application US/08420235B
; Patent No. 5801042
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
; TITLE OF INVENTION: SEQUENCES AND USES THEREOF
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/420,235B
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 45185-B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20710 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: N
; ANTI-SENSE: N
US-08-420-235B-1

Query Match 52.0%; Score 13; DB 1; Length 20710;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gcatcgccagcat 13
|||||
Db 2137 GCATCGCCAGCAT 2125

RESULT 22
US-08-793-624-1/c
; Sequence 1, Application US/08793624C
; Patent No. 6150093
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: Unique Associated Kaposi's Sarcoma Virus Sequences And
; TITLE OF INVENTION: Uses Thereof
; FILE REFERENCE: 45185-C-PCT-US/JPW
; CURRENT APPLICATION NUMBER: US/08/793,624C
; CURRENT FILING DATE: 1997-02-18
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO: 1
; LENGTH: 20710
; TYPE: DNA
; ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-08-793-624-1

Query Match 52.0%; Score 13; DB 3; Length 20710;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gcatcgccagcat 13
|||||
Db 2137 GCATCGCCAGCAT 2125

RESULT 23
PCT-US95-10194-1/c
; Sequence 1, Application PC/TUS9510194
; GENERAL INFORMATION:
; APPLICANT: The Trustees of Columbia University in the City of New York
; APPLICANT: City
; TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
; TITLE OF INVENTION: SEQUENCES AND USES THEREOF

```
;
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCI/US95/10194
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 45185-C-PCT/JPW/MSC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20710 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: N
; ANTI-SENSE: N
; PCT-US95-10194-1

Query Match 52.0%; Score 13; DB 5; Length 20710;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcatgcgcagcat 13
|||||
Db 2137 GCATGCCAGCAT 2125

RESULT 24
US-08-770-379-18/c
; Sequence 18, Application US/08770379
; Patent No. 5849564
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A.
; APPLICANT: Russo, James J.
; APPLICANT: Edelman, Isidore S.
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: POLYPEPTIDES FROM KAPOSI'S SARCOMA-ASSOCIATED
; HERPESVIRUS, DNA ENCODING SAME AND USES THEREOF
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/770,379
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 45185-F
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 35100 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-757-669A-18
```

```
;
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 52342
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 35100 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-770-379-18

Query Match 52.0%; Score 13; DB 2; Length 35100;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcatgcgcagcat 13
|||||
Db 2057 GCATGCCAGCAT 2045

RESULT 25
US-08-757-669A-18/c
; Sequence 18, Application US/08757669A
; Patent No. 6183751
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A.
; APPLICANT: Russo, James J.
; APPLICANT: Edelman, Isidore S.
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/757,669A
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 45185-F
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 35100 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-757-669A-18
```

Query Match 52.0%; Score 13; DB 4; Length 35100;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gcatgccagcat 13
|||||
DB 2057 GCATGCCAGCAT 2045

RESULT 26
US-08-851-843A-92/C
; Sequence 92, Application US/08851843A
; Patent No. 6093809
; GENERAL INFORMATION:
; APPLICANT: Cech, Thomas R.
; APPLICANT: Lingner, Joachim
; APPLICANT: Nakamura, Toru
; APPLICANT: Chapman, Karen B.
; APPLICANT: Morin, Gregg B.
; APPLICANT: Harley, Calvin B.
; APPLICANT: Andrews, William H.
; TITLE OF INVENTION: No. 6093809el Telomerase
; NUMBER OF SEQUENCES: 225
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/851.843A
; FILING DATE: 06-MAY-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/846,017
; FILING DATE: 25-APR-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/844,419
; FILING DATE: 18-APR-1997
; CLASSIFICATION:
; APPLICATION DATA:
; APPLICATION NUMBER: US 08/724,643
; FILING DATE: 01-OCT-1996
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Apple, Randolph T.
; REGISTRATION NUMBER: 36,429
; REFERENCE/DOCKET NUMBER: 015389-00293005
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 92:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 19 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
US-08-851-843A-92

Query Match 48.0%; Score 12; DB 3; Length 19;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 catgccagcat 13

DB 19 CATGCCAGCAT 8

RESULT 27
US-08-974-549A-384/C
; Sequence 384, Application US/08974549A
; Patent No. 6166178
; GENERAL INFORMATION:
; APPLICANT: Cech, Thomas R.
; APPLICANT: Lingner, Joachim
; APPLICANT: Nakamura, Toru
; APPLICANT: Chapman, Karen B.
; APPLICANT: Morin, Gregg B.
; APPLICANT: Harley, Calvin B.
; APPLICANT: Andrews, William H.
; TITLE OF INVENTION: Human Telomerase Catalytic Subunit
; NUMBER OF SEQUENCES: 727
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/974,549A
; FILING DATE: 19-NOV-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/724,643
; FILING DATE: 01-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/844,419
; FILING DATE: 18-APR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/846,017
; FILING DATE: 25-APR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/851,843
; FILING DATE: 06-MAY-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/854,050
; FILING DATE: 09-MAY-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/911,312
; FILING DATE: 14-AUG-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/912,951
; FILING DATE: 14-AUG-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/915,503
; FILING DATE: 14-AUG-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US97/17618
; FILING DATE: 01-OCT-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US97/17885
; FILING DATE: 01-OCT-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Apple, Randolph T.
; REGISTRATION NUMBER: 36,429
; REFERENCE/DOCKET NUMBER: 015389-00261005
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 384:
; SEQUENCE CHARACTERISTICS:

```

; LENGTH: 19 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..19
; OTHER INFORMATION: /note= "Tcpl.6 primer"
US-08-974-549A-384

```

```

Query Match      48.0%; Score 12; DB 4; Length 19;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 2 catcgccagcat 13
Db 19 CATGCCAGCAT 8

```

```

RESULT 28
US-08-373-124A-1692/c
; Sequence 1692, Application US/08373124A
; Patent No. 5646042
; GENERAL INFORMATION:
; APPLICANT: Stinchcomb, Dan T.
; APPLICANT: Draper, Kenneth
; APPLICANT: McSwiggen, James
; APPLICANT: Jarvis, Thale
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR
; TITLE OF INVENTION: TREATMENT OF RESTENOSIS AND
; TITLE OF INVENTION: CANCER USING RIBOZYMES
; NUMBER OF SEQUENCES: 2627
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071

```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/373,124A
; FILING DATE: January 13, 1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/245,466
; FILING DATE: May 18, 1994
; APPLICATION NUMBER: 08/192,943
; FILING DATE: February 7, 1994
; APPLICATION NUMBER: 07/987,132
; FILING DATE: December 7, 1992
; APPLICATION NUMBER: 07/936,422
; FILING DATE: August 26, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 209/035
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 1692:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 38 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single

```

```

; TOPOLOGY: linear
US-08-373-124A-1692

Query Match      48.0%; Score 12; DB 1; Length 38;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 6 gccagcatttcg 17
Db 38 GCCAGCATTTTCG 27

```

```

RESULT 29
US-08-435-628-1692/c
; Sequence 1692, Application US/08435628
; Patent No. 5817796
; GENERAL INFORMATION:
; APPLICANT: Stinchcomb, Dan T.
; APPLICANT: Draper, Kenneth
; APPLICANT: McSwiggen, James
; APPLICANT: Jarvis, Thale
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR
; TITLE OF INVENTION: TREATMENT OF RESTENOSIS AND
; TITLE OF INVENTION: CANCER USING RIBOZYMES
; NUMBER OF SEQUENCES: 2627
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/435,628
; FILING DATE: 05-MAY-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/373,124
; FILING DATE: January 13, 1995
; APPLICATION NUMBER: 08/245,466
; FILING DATE: May 18, 1994
; APPLICATION NUMBER: 08/192,943
; FILING DATE: February 7, 1994
; APPLICATION NUMBER: 07/987,132
; FILING DATE: December 7, 1992
; APPLICATION NUMBER: 07/936,422
; FILING DATE: August 26, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 209/035
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 1692:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 38 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-435-628-1692

```

```

Query Match      48.0%; Score 12; DB 1; Length 38;

```

```

Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 gccagcattcg 17
   |||||
Db 38 GCCAGCATTCG 27

RESULT 30
US-08-743-637B-175/c
; Sequence 175, Application US/08743637B
; Patent No. 5994066
; GENERAL INFORMATION:
; APPLICANT: BERGERON, Michel G.
; APPLICANT: PICARD, Francois J.
; APPLICANT: OUELLETTE, Marc
; APPLICANT: ROY, Paul H.
; TITLE OF INVENTION: SPECIES-SPECIFIC AND UNIVERSAL DNA
; TITLE OF INVENTION: PROBES AND AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND
; TITLE OF INVENTION: IDENTIFY COMMON BACTERIAL PATHOGENS AND ASSOCIATED
; TITLE OF INVENTION: ANTIBIOTIC RESISTANCE GENES FROM CLINICAL SPECIMENS ...
; NUMBER OF SEQUENCES: 273
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: QUARLES & BRADY
; STREET: 411 EAST WISCONSIN AVENUE
; CITY: MILWAUKEE
; STATE: WISCONSIN
; COUNTRY: USA
; ZIP: 53202-4497
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/743.637B
; FILING DATE: 04-NOV-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/526,840
; FILING DATE: 11-SEP-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: BAKER, Jean C.
; REGISTRATION NUMBER: 35,433
; REFERENCE/DOCKET NUMBER: 850586.90012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (414) 277-5591
; INFORMATION FOR SEQ ID NO: 175:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 660 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-743-637B-175

Query Match 48.0%; Score 12; DB 2; Length 660;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 gccagcattcg 19
   |||||
Db 627 CAGCATTCGAT 616

RESULT 31
US-08-840B-175/c
; Sequence 175, Application US/08526840B
; Patent No. 6001564
; GENERAL INFORMATION:
; APPLICANT: BERGERON, Michel G.

```

```

; APPLICANT: OUELLETTE, Marc
; APPLICANT: ROY, Paul H.
; TITLE OF INVENTION: SPECIFIC AND UNIVERSAL PROBES AND
; TITLE OF INVENTION: AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND IDENTIFY
; TITLE OF INVENTION: COMMON BACTERIAL PATHOGENS AND ANTIBIOTIC RESISTANCE
; TITLE OF INVENTION: FROM CLINICAL SPECIMENS FOR ROUTINE DIAGNOSIS IN ...
; NUMBER OF SEQUENCES: 177
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: QUARLES & BRADY
; STREET: 411 East Wisconsin Avenue
; CITY: Milwaukee
; STATE: Wisconsin
; COUNTRY: USA
; ZIP: 53202-4497
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/526,840B
; FILING DATE: 11-SEP-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/304,732
; FILING DATE: 12-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BAKER, Jean C.
; REGISTRATION NUMBER: 35,433
; REFERENCE/DOCKET NUMBER: 850586.90012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (414) 277-5000
; TELEFAX: (414) 277-5591
; INFORMATION FOR SEQ ID NO: 175:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 660 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-526-840B-175

Query Match 48.0%; Score 12; DB 3; Length 660;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 gccagcattcg 19
   |||||
Db 627 CAGCATTCGAT 616

RESULT 32
US-08-998-416-767
; Sequence 767, Application US/08998416
; Patent No. 6239264
; GENERAL INFORMATION:
; APPLICANT: Philippsen, Peter
; APPLICANT: Pohlmann, Rainer
; APPLICANT: Steiner, Sabine
; APPLICANT: Mohr, Christine
; APPLICANT: Wendland, Jurgen
; APPLICANT: Knechtle, Philipp
; APPLICANT: Reibischung, Corinne
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPHII
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 1152
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6239264artis Corporation
; STREET: 3054 Cornwallis Road
; CITY: Research Triangle Park
; STATE: No. 6239264th Carolina
; COUNTRY: USA

```


ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/998,416
FILING DATE: 24-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meligs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 767:
SEQUENCE CHARACTERISTICS:
LENGTH: 715 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: PAG1495RP
US-08-998-416-767

Query Match 48.0%; Score 12; DB 4; Length 715;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 tcgccagcattt 15
Db 356 TCGCCAGCAATT 367

RESULT 33
US-08-927-128-3
Sequence 3, Application US/08927128
Patent No. 6127150
GENERAL INFORMATION:
APPLICANT: Coolidge, Thomas
APPLICANT: Wagner, Fred
APPLICANT: ven Heeke, Gino
APPLICANT: Schuster, Sheldon
APPLICANT: Stout, Jay
APPLICANT: Wylie, Dwane
TITLE OF INVENTION: PURIFICATION DIRECTED CLOSING OF PEPTIDES
NUMBER OF SEQUENCES: 28
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merchants & Gould
STREET: 3100 No. 6127150 West Center, 90 S. 7th Street
CITY: Minneapolis
STATE: MN
COUNTRY: U.S.A.
ZIP: 55402
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/927,128
FILING DATE: 05-SEP-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/680,004
FILING DATE: 15-JUL-1995

ATTORNEY/AGENT INFORMATION:
NAME: Carter, Charles G
REGISTRATION NUMBER: 35,093
REFERENCE/DOCKET NUMBER: 8648.2USDI
TELECOMMUNICATION INFORMATION:
TELEPHONE: 612/332-5300
TELEFAX: 612/332-9081
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 780 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE:
ORIGINAL SOURCE:
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 1..780
OTHER INFORMATION:
US-08-927-128-3

Query Match 48.0%; Score 12; DB 3; Length 780;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 14 ttgcgtctcgt 25
Db 532 TTGCATCTCTGT 543

RESULT 34
US-08-990-823-39
Sequence 39, Application US/08990823D
Patent No. 6228371
GENERAL INFORMATION:
APPLICANT: Nano, Francis
TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
TITLE OF INVENTION: Immunostimulatory Peptides
FILE REFERENCE: 49086
CURRENT APPLICATION NUMBER: US/08/990,823D
CURRENT FILING DATE: 1997-12-15
EARLIER APPLICATION NUMBER: US 96/10375
EARLIER FILING DATE: 1996-06-14
EARLIER APPLICATION NUMBER: 60/000,254
EARLIER FILING DATE: 1995-06-15
NUMBER OF SEQ ID NOS: 113
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 39
LENGTH: 841
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
FEATURE:
NAME/KEY: Modified base
OTHER INFORMATION: n represents a or g or c or t/u
US-08-990-823-39

Query Match 48.0%; Score 12; DB 4; Length 841;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gcctcgccagca 12
Db 370 gcctcgccagca 381

RESULT 35
US-08-595-868C-11

```

; Sequence 11, Application US/08595868C
; Patent No. 5962270
; GENERAL INFORMATION:
; APPLICANT: Wagner, Fred
; APPLICANT: Stout, Jay
; APPLICANT: Henriksen, Dennis
; APPLICANT: Partridge, Bruce
; APPLICANT: Holmquist, Bart
; APPLICANT: Frank, Julie
; TITLE OF INVENTION: RECOMBINANT PREPARATION OF CALCITONIN FRAGMENTS AND US
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merchant & Gould
; STREET: 3100 No. 5962270 West Center, 90 S. 7th Street
; CITY: Minneapolis
; STATE: MN
; COUNTRY: U.S.A.
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/595,868C
; FILING DATE: 06-FEB-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Carter, Charles G
; REGISTRATION NUMBER: 35,093
; REFERENCE/DOCKET NUMBER: 8648.590S01
; TELEPHONE: 612/332-5300
; TELEFAX: 612/332-9081
; TELEX:
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 864 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; FEATURE:
; NAME/KEY: Coding Sequence
; LOCATION: 1..864
; OTHER INFORMATION:
; US-08-595-868C-11

```

```

Query Match 48.0%; Score 12; DB 2; Length 864;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 14 ttcgatctctgt 25
| | | | | | | | | |
Db 532 TTCGATCTCTGT 543

```

```

RESULT 36
US-09-139-819A-11
; Sequence 11, Application US/09139819A
; Patent No. 6251635
; GENERAL INFORMATION:
; APPLICANT: WAGNER, Fred W.
; APPLICANT: STOUT, Jay S.
; APPLICANT: HENRIKSEN, Dennis B.

```

```

; APPLICANT: PARTRIDGE, Bruce E.
; APPLICANT: HOLMQUIST, Bart
; APPLICANT: FRANK, Julie A.
; TITLE OF INVENTION: RECOMBINANT PREPARATION OF CALCITONIN
; TITLE OF INVENTION: FRAGMENTS AND USE THEREOF IN THE PREPARATION OF CALCITONIN
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY & LARDNER
; STREET: 3000 K Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20007-5109
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/139,819A
; FILING DATE: 25-AUG-1998
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/595,868
; FILING DATE: 06-FEB-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 089187/0144
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 867 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..864
; US-09-139-819A-11

```

```

Query Match 48.0%; Score 12; DB 4; Length 867;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 14 ttcgatctctgt 25
| | | | | | | | | |
Db 532 TTCGATCTCTGT 543

```

```

RESULT 37
US-08-624-125-9/C
; Sequence 9, Application US/08624125
; Patent No. 5744341
; GENERAL INFORMATION:
; APPLICANT: CUNNINGHAM JR., FRANCIS X.
; APPLICANT: SUN, ZAIREN
; TITLE OF INVENTION: GENES OF CAROTENOID BIOSYNTHESIS AND
; TITLE OF INVENTION: METABOLISM AND A SYSTEM FOR SCREENING SUCH GENES
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MATER & NEUPSTADT,
; ADDRESS: P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/624,125
FILING DATE: 29-MAR-1996
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: KELBER, STEVEN B.
REGISTRATION NUMBER: 30,073
REFERENCE/DOCKET NUMBER: 2747-063-27
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-413-3000
TELEFAX: 703-413-2220
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 954 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-624-125-9

Query Match 48.0%; Score 12; DB 1; Length 954;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 catgccagcatt 13
|||||
DB 663 CATGCCAGCATT 652

RESULT 38
US-08-781-562-2/c
Sequence 2, Application US/08/81562
Patent No. 5763589
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Goli, Surya K.
TITLE OF INVENTION: NOVEL HUMAN MEMBRANE PROTEIN
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/781,562
FILING DATE: Herewith
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0181 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:

LENGTH: 1188 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: Consensus
CLONE: Consensus
US-08-781-562-2

Query Match 48.0%; Score 12; DB 1; Length 1188;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 cgcgcagcatttc 16
|||||
DB 72 CGCCAGCATTTC 61

RESULT 39
US-09-064-693A-20/c
Sequence 20, Application US/09064693A
Patent No. 6210937
GENERAL INFORMATION:
APPLICANT: Ward, Thomas E.
TITLE OF INVENTION: DEVELOPMENT OF GENETICALLY
ENGINEERED BACTERIA FOR PRODUCTION
OF A SPECIFIC PLASTICS PRECURSOR
TITLE OF INVENTION: OF A SPECIFIC PLASTICS PRECURSOR
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: W. Gary Goodson
ADDRESSEE: INEL--Lockheed Martin Idaho
ADDRESSEE: Technologies Co.
STREET: P.O. Box 1625
CITY: Idaho Falls
STATE: Idaho
COUNTRY: USA
ZIP: 83415-3810
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage
COMPUTER: Toshiba Satellite Pro T2150CDS
OPERATING SYSTEM: Windows95
SOFTWARE: Word Perfect 7.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/064,693A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: W. Gary Goodson
REGISTRATION NUMBER: 22,387
REFERENCE/DOCKET NUMBER: LIT-PI-296
TELECOMMUNICATION INFORMATION:
TELEPHONE: (208)526-9469
TELEFAX: (208)526-8339
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 1239 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-09-064-693A-20

Query Match 48.0%; Score 12; DB 4; Length 1239;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 atccgcagcatt 14
|||||
DB 1137 ATCCGCCAGCATT 1126

```

RESULT 40
US-08-985-950-1
; Sequence 1, Application US/08985950
; Patent No. 6140076
; GENERAL INFORMATION:
; APPLICANT: Adema, Gosse Jan
; TITLE OF INVENTION: Isolated Mammalian Monocyte Cell Genes;
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DNAX Research Institute
; STREET: 901 California Avenue
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/985,950
; FILING DATE: 05-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/041,279
; FILING DATE: 21-MARCH-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/033,181
; FILING DATE: 16-DEC-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/032,252
; FILING DATE: 06-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Ching, Edwin P.
; REGISTRATION NUMBER: 34,090
; REFERENCE/DOCKET NUMBER: DX0670K
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650)852-9196
; TELEFAX: (650)496-1204
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1249 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 154..1062
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 211..1062
US-08-985-950-1

```

```

Query Match 48.0%; Score 12; DB 3; Length 1249;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY - 5 cgcacgatttc 16
| | | | |
Db 197 CGCCAGCATTTC 208

```

```

RESULT 41
US-08-484-575A-19/c
; Sequence 19, Application US/08484575A
; Patent No. 5925358
; GENERAL INFORMATION:
; APPLICANT: Mark D. Cochran and David E. Junker

```

```

; TITLE OF INVENTION: Recombinant Fowlpox Viruses and Uses Thereof
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John P. White
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/484,575A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: White Esq, John P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)278-0450
; TELEFAX: (212)391-0525
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1305 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1305
US-08-484-575A-19

```

```

Query Match 48.0%; Score 12; DB 2; Length 1305;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ccagcatttcga 18
| | | | |
Db 780 CCAGCATTTCGA 769

```

```

RESULT 42
US-08-477-459-19/c
; Sequence 19, Application US/08477459
; Patent No. 6001369
; GENERAL INFORMATION:
; APPLICANT: Mark D. Cochran
; TITLE OF INVENTION: Recombinant Fowlpox Viruses and Uses
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John P. White
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,459
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:

```

```

; NAME: White Esq, John P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1305 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1305
US-08-477-459-19

Query Match 48.0%; Score 12; DB 3; Length 1305;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ccagcatttcga 18
|||||
Db 780 CCAGCATTTCGA 769

RESULT 44
US-08-486-414-19/c
; Sequence 19, Application US/08486414B
; Patent No. 6136318
; GENERAL INFORMATION:
; APPLICANT: Cochran, Mark D.
; APPLICANT: Junker, David E.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUSES AND USES THEREOF
; FILE REFERENCE: 42771D
; CURRENT APPLICATION NUMBER: US/08/486,414B
; CURRENT FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 19
; LENGTH: 1305
; TYPE: DNA
; ORGANISM: Fowlpox virus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(432)
US-08-486-414-19

Query Match 48.0%; Score 12; DB 3; Length 1305;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ccagcatttcga 18
|||||
Db 780 CCAGCATTTCGA 769

RESULT 45
PCT-US94-01826A-19/c
; Sequence 19, Application PC/TUS9401826A
; GENERAL INFORMATION:
; APPLICANT: Syntro Corporation, et al.
; TITLE OF INVENTION: Recombinant Fowlpox Virus S-FPV-043 and Uses Thereof
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: John P. White
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/479,869
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION NUMBER: 08/024,156
; FILING DATE: 26-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: White Esq, John P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)977-9550
; TELEFAX: (212)664-0525
; TELEX: 422523
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1305 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS

```

```
; TELEX: 422523
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1305 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1305
PCT-US94-01826A-19
```

```
Query Match      48.0%; Score 12; DB 5; Length 1305;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      7 ccagcatttcga 18
        |||
Db      780 CCAGCATTTCGA 769
```

Search completed: October 9, 2001, 15:55:39
Job time: 13215 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 9, 2001, 15:15:32 ; Search time 5323.87 Seconds
(without alignments)
44.389 Million cell updates/sec

Title: us-09-396-196f-7
Perfect score: 25
Sequence: 1 gcatgccagcatttcgatccctcgt 25

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 10228115 seqs, 4726426750 residues

Word size : 9

Total number of hits satisfying chosen parameters: 367121

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : EST.*
1: gb_est1.*
2: gb_est2.*
3: gb_est3.*
4: gb_est4.*
5: gb_est5.*
6: gb_est6.*
7: gb_est7.*
8: gb_est8.*
9: gb_est9.*
10: gb_est10.*
11: gb_est11.*
12: gb_est12.*
13: gb_est13.*
14: gb_est14.*
15: gb_est15.*
16: gb_est16.*
17: gb_est17.*
18: gb_est18.*
19: gb_est19.*
20: gb_est20.*
21: gb_est21.*
22: gb_est22.*
23: gb_est23.*
24: gb_est24.*
25: gb_est33.*
26: gb_est34.*
27: gb_est35.*
28: gb_est36.*
29: gb_est37.*
30: gb_est38.*
31: gb_est39.*
32: gb_est40.*
33: em_estba.*
34: em_estfun.*
35: em_esthum1.*
36: em_esthum2.*
37: em_esthum3.*
38: em_esthum4.*
39: em_esthum5.*
40: em_esthum6.*
41: em_esthum7.*
42: em_esthum8.*
43: em_esthum9.*

44: em_esthum10.*
45: em_esthum11.*
46: em_esthum12.*
47: em_esthum13.*
48: em_esthum14.*
49: em_esthum15.*
50: em_esthum16.*
51: em_esthum17.*
52: em_esthum18.*
53: em_esthum19.*
54: em_esthum20.*
55: em_esthum21.*
56: em_esthum22.*
57: em_esthum23.*
58: em_esthum24.*
59: em_esthum25.*
60: em_esthum26.*
61: em_esthum27.*
62: em_esthum28.*
63: em_estin1.*
64: em_estin2.*
65: em_estin3.*
66: em_estin4.*
67: em_estin5.*
68: em_estom1.*
69: em_estom2.*
70: em_estov1.*
71: em_estov2.*
72: em_estpl1.*
73: em_estpl2.*
74: em_estpl3.*
75: em_estpl4.*
76: em_estpl5.*
77: em_estpl6.*
78: em_estpl7.*
79: em_estpl8.*
80: em_estpl9.*
81: em_estpl10.*
82: em_estro1.*
83: em_estro2.*
84: em_estro3.*
85: em_estro4.*
86: em_estro5.*
87: em_estro6.*
88: em_estro7.*
89: em_estro8.*
90: em_estro9.*
91: em_estro10.*
92: em_estro11.*
93: em_estro12.*
94: em_estro13.*
95: em_estro14.*
96: em_estro15.*
97: em_estro16.*
98: em_estro17.*
99: em_estro18.*
100: em_estro19.*
101: em_estro20.*
102: gb_est25.*
103: gb_est26.*
104: gb_est27.*
105: gb_est28.*
106: gb_est29.*
107: gb_est30.*
108: gb_est31.*
109: gb_est32.*
110: gb_est41.*
111: gb_est42.*
112: gb_est43.*
113: gb_est44.*
114: gb_est45.*
115: gb_est46.*
116: gb_est47.*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

190: gb_est121:*
191: gb_est122:*
192: gb_est123:*
193: gb_est124:*
194: gb_est125:*
195: gb_est126:*
196: gb_est127:*
197: gb_est128:*
198: gb_est129:*
199: gb_est130:*
200: gb_est131:*
201: gb_est132:*
202: gb_est133:*
203: gb_est134:*
204: gb_est135:*
205: gb_est136:*
206: gb_est137:*
207: gb_est138:*
208: gb_est139:*
209: gb_est140:*
210: gb_est141:*
211: gb_est142:*
212: gb_est143:*
213: gb_est144:*
214: gb_est145:*
215: gb_est146:*
216: gb_est147:*
217: gb_est148:*
218: gb_est149:*
219: gb_est150:*
220: gb_est151:*
221: gb_est152:*
222: gb_est153:*
223: gb_est154:*
224: gb_est155:*
225: gb_est156:*
226: gb_est157:*
227: gb_est158:*
228: gb_est159:*
229: gb_est160:*
230: gb_est161:*
231: gb_est162:*
232: gb_est163:*
233: gb_est164:*
234: gb_est165:*
235: gb_est166:*
236: gb_est167:*
237: gb_est168:*
238: gb_est169:*
239: gb_est170:*
240: gb_est171:*
241: gb_est172:*
242: gb_est173:*
243: gb_est174:*
244: gb_est175:*
245: gb_est176:*
246: gb_est177:*
247: gb_est178:*
248: gb_est179:*
249: gb_est180:*
250: gb_est181:*
251: gb_est182:*
252: gb_est183:*
253: gb_est184:*
254: gb_est185:*
255: gb_est186:*
256: gb_est187:*
257: gb_est188:*
258: gb_est189:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	17	68.0	498	21	AI532390
2	16	64.0	429	107	AU030894
3	16	64.0	510	107	AU031475
4	16	64.0	666	107	AU030895
5	16	64.0	704	107	AU078750
6	15	60.0	258	246	AZ578088
7	15	60.0	338	162	BE065518
8	15	60.0	341	31	AV551712
9	15	60.0	393	158	H60965
10	15	60.0	394	189	T86053
11	15	60.0	394	190	W23358
12	15	60.0	399	162	BE044431
13	15	60.0	414	18	AI268648
14	15	60.0	435	223	AQ027122
15	15	60.0	437	158	H69200
16	15	60.0	438	187	N92081
17	15	60.0	445	32	AV722315
18	15	60.0	466	1	AA026650
19	15	60.0	475	158	H60979
20	15	60.0	491	23	AI692369
21	15	60.0	491	146	BF211097
22	15	60.0	492	251	AZ924681
23	15	60.0	507	189	T84279
24	15	60.0	528	158	H74120
25	15	60.0	539	170	BF839030
26	15	60.0	547	138	AW588900
27	15	60.0	581	7	AA405993
28	15	60.0	622	137	BE557935
29	15	60.0	630	137	BE583417
30	15	60.0	684	145	BF156891
31	15	60.0	1096	145	BF144032
32	14	56.0	195	189	W06020
33	14	56.0	200	251	AZ922752
34	14	56.0	203	111	AW065325
35	14	56.0	203	155	BG589383
36	14	56.0	205	155	BG558891
37	14	56.0	211	235	AQ931496
38	14	56.0	231	18	AI329089
39	14	56.0	242	103	AI902416
40	14	56.0	251	125	BB090429
41	14	56.0	271	115	AW423795
42	14	56.0	286	3	AA165335
43	14	56.0	311	128	BB216989
44	14	56.0	322	152	BG354207
45	14	56.0	322	168	BF714260

ALIGNMENTS

RESULT	1
AI532390/c	
LOCUS	
DEFINITION	AI532390.1 GI:4446525
ACCESSION	AI532390
VERSION	EST
KEYWORDS	fruit fly.
SOURCE	Drosophila melanogaster
ORGANISM	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE	1 (bases 1 to 498)
AUTHORS	Harvey,D., Hong,L., Evans-Holm,M., Pendleton,J., Su,C., Brokstein ,P., Lewis,S. and Rubin,G.M.

TITLE
JOURNAL
COMMENT

BDGP/HHMI Drosophila EST Project
Unpublished (1997)
Contact: Harvey, D.
G. M. Rubin-Molecular and Cell Biology
University of California Berkeley
539 LSA, Berkeley, CA 94720-3200, USA
Fax: 510 643 9947
Email: <http://www.fruitfly.org/EST>, est@fruitfly.berkeley.edu

Plate: 38 row: H column: 7
High quality sequence stop: 433.
Location/Qualifiers

FEATURES
source

1..498
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="SP03891"
/clone_lib="Sp Drosophila melanogaster Schneider L2 cell culture pot2"
/lab_host="DH5-alpha"
/note="Vector: pOT2; Site_1: EcoRI; Site_2: XhoI; Sized fractionated cDNAs were directly ligated into pOT2.
Plasmid cDNA library."
BASE COUNT 158 a 106 c 124 g 110 t
ORIGIN

Query Match 68.0% Score 17: DB 21: Length 498;
Best Local Similarity 100.0%; Pred. No. 2.5; 0; Indels 0; Gaps 0;
Matches 17; Conservative 0; Mismatches 0;

Qy 1 gcatcgccagcatttcg 17
|||||
Db 266 GCATCGCCAGCATTCG 250

RESULT 2

LOCUS

AI532390.1 429 bp mRNA EST 29-OCT-1998
DEFINITION AU030894 Rice cDNA from immature leaf including apical meristem
Oryza sativa cDNA clone E60405_1A, mRNA sequence.

ACCESSION AU030894

VERSION AU030894.1

KEYWORDS GI:3766784

SOURCE Oryza sativa

ORGANISM Oryza sativa

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Ehrhartoideae; Oryzaceae; Oryza.

Sasaki,T. and Yamamoto,K.

Rice cDNA from immature leaf including apical meristem

Unpublished (1997)

CONTACT Takuji Sasaki

National Institute of Agrobiological Resources

Rice Genome Research Program

2-1-2 Kannondai,Tsukuba

Ibaraki,

Japan 305

Tel: 0298-38-7441

Fax: 0298-38-7468

Email: tsasaki@abr.affrc.go.jp

PROJECT = 'RGP'

Location/Qualifiers

1..429

/organism="Oryza sativa"

/strain="Nipponbare"

/db_xref="taxon:4530"

/clone="E60405_1A"

/clone_lib="Rice cDNA from immature leaf including apical meristem"

/dev_stage="immature"

/note="Organ: leaf; immature leaf including apical meristem (under long day condition)"

BASE COUNT 106 a 122 c 101 g 99 t 1 others

ORIGIN

Query Match 64.0%; Score 16; DB 107; Length 429;
 Best Local Similarity 100.0%; Pred. No. 10;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 catgccagcattcg 17
 |||||

Db 302 CATGCCAGCATTTCG 287

RESULT 3

AU031475/c 510 bp mRNA EST 30-OCT-1998

LOCUS AU031475 Rice cDNA from immature leaf including apical meristem
 DEFINITION Oriza sativa cDNA clone E61659_22, mRNA sequence.

ACCESSION AU031475

VERSION AU031475.1 GI:3767365

KEYWORDS EST.

SOURCE Oriza sativa.

ORGANISM Oriza sativa

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzeae; Oryza.

REFERENCE 1 (bases 1 to 510)

AUTHORS Sasaki,T. and Yamamoto,K.

TITLE Rice cDNA from immature leaf including apical meristem

JOURNAL Unpublished (1997)

COMMENT Contact: Takuji Sasaki

National Institute of Agrobiological Resources

Rice Genome Research Program

2-1-2 Kannondai,Tsukuba

Ibaraki,

Japan 305

Tel: 0298-38-7441

Fax: 0298-38-7468

Email: tsasaki@abr.affrc.go.jp

PROJECT = "RGP".

Location/Qualifiers

1..510

/organism="Oryza sativa"

/strain="Nipponbare"

/db_xref="taxon:4530"

/clone_lib="Rice cDNA from immature leaf including apical

meristem"

/dev_stage="immature"

/note="Organ: leaf; immature leaf including apical

meristem (under long day condition)"

176 a 98 c 99 g 137 t

BASE COUNT

ORIGIN

Query Match 64.0%; Score 16; DB 107; Length 510;
 Best Local Similarity 100.0%; Pred. No. 10;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 catgccagcattcg 17
 |||||

Db 136 CATGCCAGCATTTCG 121

RESULT 4

AU030895/c 666 bp mRNA EST 29-OCT-1998

LOCUS AU030895 Rice cDNA from immature leaf including apical meristem

DEFINITION Oriza sativa cDNA clone E60405_22, mRNA sequence.

ACCESSION AU030895

VERSION AU030895.1 GI:3766785

KEYWORDS EST.

SOURCE Oriza sativa.

ORGANISM Oriza sativa

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzeae; Oryza.

1 (bases 1 to 666)

AUTHORS Sasaki,T. and Yamamoto,K.

TITLE Rice cDNA from immature leaf including apical meristem

JOURNAL Unpublished (1997)

COMMENT Contact: Takuji Sasaki

National Institute of Agrobiological Resources

Rice Genome Research Program

2-1-2 Kannondai,Tsukuba

Ibaraki,

Japan 305

Tel: 0298-38-7441

Fax: 0298-38-7468

Email: tsasaki@abr.affrc.go.jp

PROJECT = "RGP".

Location/Qualifiers

1..666

/organism="Oryza sativa"

/strain="Nipponbare"

/db_xref="taxon:4530"

/clone_lib="Rice cDNA from immature leaf including apical

meristem"

/dev_stage="immature"

/note="Organ: leaf; immature leaf including apical

meristem (under long day condition)"

195 a 149 c 131 g 181 t

10 others

BASE COUNT

ORIGIN

Query Match 64.0%; Score 16; DB 107; Length 666;
 Best Local Similarity 100.0%; Pred. No. 10;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 catgccagcattcg 17
 |||||

Db 293 CATGCCAGCATTTCG 278

RESULT 5

AU078750/c

LOCUS AU078750 Rice panicle at flowering stage Oryza sativa cDNA clone

DEFINITION E0723_42, mRNA sequence.

ACCESSION AU078750

VERSION AU078750.1 GI:6062509

KEYWORDS EST.

SOURCE Oryza sativa.

ORGANISM Oryza sativa

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

Ehrhartoideae; Oryzeae; Oryza.

1 (bases 1 to 704)

AUTHORS Sasaki,T. and Yamamoto,K.

TITLE Rice cDNA from panicle at flowering stage

JOURNAL Unpublished (1996)

COMMENT Contact: Takuji Sasaki

National Institute of Agrobiological Resources

Rice Genome Research Program

2-1-2 Kannondai,Tsukuba

Ibaraki,

Japan 305

Tel: 0298-38-7441

Fax: 0298-38-7468

Email: tsasaki@abr.affrc.go.jp

PROJECT = "RGP".

Location/Qualifiers

1..704

/organism="Oryza sativa"

/strain="Nipponbare"

/db_xref="taxon:4530"

BASE COUNT 203 a 167 c 142 g 191 t 1 others
 ORIGIN
 /clone="E0723_42"
 /clone_lib="Rice panicle at flowering stage"
 /dev_stage="flowering stage"
 /note="Organ: panicle; Rice cDNA from panicle at flowering stage"

Query Match 64.0%; Score 16; DB 107; Length 704;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 catgccagcatttcg 17
 |||||

Db 346 CATGCCAGCATTCG 331

RESULT 6
 AZ578088/c 258 bp DNA GSS 08-DEC-2000
 LOCUS
 DEFINITION 18f02 Shot-gun genomic library of Rhizobium strain ANU265 Rhizobium
 sp. NGR234 genomic clone 18f02, DNA sequence.

ACCESSION AZ578088
 VERSION AZ578088.1 GI:11605015
 KEYWORDS GSS.

ORGANISM Rhizobium sp. NGR234.
 Rhizobium sp. NGR234
 Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 Rhizobiaceae; Rhizobium.

REFERENCE 1 (bases 1 to 258)
 AUTHORS Viprey,V., Rosenthal,A., Broughton,W.J. and Perret,X.
 TITLE Genetic snapshots of the Rhizobium species NGR234 genome
 JOURNAL Genome Biology.com 1 (6), 0014.1-0014.7 (2000)
 COMMENT Contact: Virginie Viprey
 Laboratoire de Biologie Moleculaire des Plantes Superieures
 University of Geneva
 1 Chemin de l'Imperatrice, Chambesey/Geneva 1292, Switzerland
 Tel: +44(0)1603450000
 Fax: +44(0)1603450045
 Email: virginie.viprey@bsrc.ac.uk
 Class: shotgun.

FEATURES
 source
 1..258
 Location/Qualifiers
 /organism="Rhizobium sp. NGR234"
 /strain="ANU265"
 /db_xref="taxon:394"
 /clone="18f02"
 /clone_lib="Shot-gun genomic library of Rhizobium strain ANU265"
 /note="Vector: M13; derivative strain of NGR234 cured of pNGR234a"

BASE COUNT 51 a 82 c 79 g 46 t
 ORIGIN

Query Match 60.0%; Score 15; DB 246; Length 258;
 Best Local Similarity 100.0%; Pred. No. 42;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 tcgccagcatttcga 18
 |||||

Db 175 TCGCCAGCATTCGA 161

RESULT 7
 BE065518/c 338 bp mRNA EST 09-JUN-2000
 LOCUS
 DEFINITION RC3-BT0316-170200-014-b06 BT0316 Homo sapiens cDNA, mRNA sequence.
 ACCESSION BE065518
 VERSION BE065518.1 GI:8410168
 KEYWORDS EST.
 SOURCE human.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
 AUTHORS

1 (bases 1 to 338)
 Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
 Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
 Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bata,G.S., Simpson,D.H.,
 Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
 ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
 Simpson,A.J.

TITLE

Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags

JOURNAL

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

MEDLINE

20202663

COMMENT

Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001

Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2=RC3-BT0316-170>)
 200-014-b066t3=2000-02-17&t4=1)

Seq primer: puc 18 forward
 High quality sequence start: 17
 High quality sequence stop: 338.

FEATURES

source

1..338
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="BT0316"
 /note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
 SmaI; A mini-library was made by cloning products derived
 from ORESTES PCR (U.S. Letters Patent application No. 196
 716 - Ludwig Institute for Cancer Research) profiles
 into the puc 18 vector. Reverse transcription of tissue
 mRNA and cDNA amplification were performed under low
 stringency conditions."

BASE COUNT 76 a 100 c 66 g 96 t

ORIGIN

Query Match 60.0%; Score 15; DB 162; Length 338;
 Best Local Similarity 100.0%; Pred. No. 42;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 catgccagcatttc 16

|||||

Db 97 CATGCCAGCATTC 83

RESULT 8

LOCUS

AV551712

DEFINITION

AV551712 Arabidopsis thaliana roots Columbia Arabidopsis thaliana

CDNA clone RZ13e05R 5', mRNA sequence.

ACCESSION

AV551712

VERSION

AV551712.1

KEYWORDS

EST.

SOURCE

thale cress.

ORGANISM

Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.

1 (bases 1 to 341)

Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.

A large scale analysis of cDNA in Arabidopsis thaliana: Generation

of 12,028 non-redundant expressed sequence tags from normalized and

size-selected cDNA libraries

JOURNAL
MEDLINE
COMMENT

DNA Res. 7, 175-180 (2000)
20363093
Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizukazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES
source

1. .341
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="RZ13e05R"
/clone_lib="Arabidopsis thaliana roots Columbia"
/tissue_type="roots"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 73 a 85 c 53 g 130 t
ORIGIN

Query Match 60.0%; Score 15; DB 31; Length 341;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 tcgcgcagcatttcca 18
|||||
Db 322 TCGCCAGCATTTCGA 336

RESULT 9
LOCUS H60965 393 bp mRNA EST 06-OCT-1995
DEFINITION Yr22c07.r1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone H60965
ACCESSION H60965.1 GI:1013797
VERSION H60965
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 393)
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston, R., Williamson, A., Wohldmann, P. and Wilson, R.
The WashU-Merck EST Project
Unpublished (1995)
Contact: Wilton RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 1078
High quality sequence stops: 299
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1078 Std Error: 0.00
Seq primer: M13RP1
High quality sequence stop: 299.
Location/Qualifiers
1. .393
/organism="Homo sapiens"
/db_xref="GDB:3775159"
/db_xref="taxon:9606"
/clone="IMAGE:206028"
/clone_lib="Soares fetal liver spleen INFLS"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"

note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site_1: Pac I; Site_2: Eco RI; 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5', AACCTGGAAGAATTAATAAGATCTTTTCTTTTCTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 88 a 112 c 75 g 114 t
ORIGIN

Query Match 60.0%; Score 15; DB 158; Length 393;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 catgcgcagcatttc 16
|||||
Db 118 CATGCCAGCATTTC 104

RESULT 10
LOCUS T86053/c 394 bp mRNA EST 17-MAR-1995
DEFINITION Yd62q05.r1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone T86053
ACCESSION T86053
VERSION T86053
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 394)
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston, R., Williamson, A., Wohldmann, P. and Wilson, R.
The WashU-Merck EST Project
Unpublished (1995)
Contact: Wilton RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 1146
High quality sequence stops: 180
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1146 Std Error: 0.00
Seq primer: M13RP1
High quality sequence stop: 180.
Location/Qualifiers
1. .394
/organism="Homo sapiens"
/db_xref="GDB:468473"
/db_xref="taxon:9606"
/clone="IMAGE:112856"
/clone_lib="Soares fetal liver spleen INFLS"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site_1: Pac I; Site_2: Eco RI; 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5', AACCTGGAAGAATTAATAAGATCTTTTCTTTTCTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."

```

BASE COUNT      86 a      98 c      82 g      122 t      6 others
ORIGIN

Query Match      60.0%; Score 15; DB 189; Length 394;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      2 catcgccagcatttc 16
      |||
Db      183 CATCGCCAGCATTTTC 169

RESULT 11
W23358
LOCUS      394 bp      mRNA      EST      12-DEC-1996
DEFINITION      SWAMCA507SK Brugia malayi adult male cDNA (SAW94NL-BmAM) Brugia
malayi cDNA clone SWAMCA507 5', mRNA sequence.
ACCESSION      W23358
VERSION      W23358.1 GI:1300212
KEYWORDS      EST.
SOURCE      Brugia malayi.
ORGANISM      Brugia malayi.
Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
Onchocercidae; Brugia.
1 (bases 1 to 394)
Williams, S.A.
Genes expressed in adult males of Brugia malayi
Unpublished (1995)
Contact: Steven A. Williams
Molecular Parasitology
Smith College Department of Biological Sciences
Department of Biological Sciences, Clark Science Center, Smith
College, Northampton, MA, 01063, USA
Tel: 4135853826
Fax: 4135853786
Email: genome@smith.edu
Seq primer: pBluescript SK.
FEATURES
      source
      1..394
      Location/Qualifiers
      /organism="Brugia malayi"
      /strain="RGS Labs"
      /db_xref="taxon:6279"
      /clone="SWAMCA507"
      /lab_host="XLI-Blue MRF"
      /note="vector: lambda UniZap XR; Site_1: EcoR I; Site_2:
      Xho I; Lymphatic filarial nematode parasite of humans.
      mRNA was prepared from adult males of Brugia malayi
      isolated from birds and converted to double stranded cDNA
      using reverse transcriptase and oligo(dT) followed by
      RNase H and DNAPol I. The library had 4.6 x 10E6
      independent recombinants and average insert size was 800
      base pairs. The library was constructed by Noelle Ling.
      The library is available from Dr. S.A. Williams, email
      genome@smith.edu."
BASE COUNT      151 a      60 c      78 g      103 t      2 others
ORIGIN

Query Match      60.0%; Score 15; DB 190; Length 394;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      6 gccagcatttcgatc 20
      |||
Db      32 GCCAGCATTTTCGATC 46

RESULT 12
BE044431
LOCUS      399 bp      mRNA      EST      08-JUN-2000
DEFINITION      ho45d04.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
IMAGE:3040327 3', mRNA sequence.
BE044431 GI:8361484
EST.
human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 399)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.

```

DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
 Seq primer: -40UP from Gibco

High quality sequence stop: 398.

FEATURES

source
 1. .414
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1910913"
 /clone_lib="NCI_CGAP_Lu5"
 /tissue_type="carcinoid"
 /lab_host="DH10B"

/note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
 modified polylinker; 1st strand cDNA was prepared from
 neuroendocrine lung carcinoid, and was then primed with a
 Not I - oligo(dT) primer. Double-stranded cDNA was ligated
 to Eco RI adaptors (Pharmacia), digested with Not I and
 cloned into the Not I and Eco RI sites of the modified
 pT7T3 vector. Library is normalized. Library was
 constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 122 a 82 c 109 g 101 t
 ORIGIN

Query Match 60.0%; Score 15; DB 18; Length 414;
 Best Local Similarity 100.0%; Pred. No. 42;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 catgccagcatttc 16
 |||||

Db 353 CATGCCAGCATTTC 367

RESULT 14
 AQ027122 435 bp DNA GSS 30-JUN-1998
 LOCUS CIT-HSP-2323H13.TF CIT-HSP Homo sapiens genomic clone 2323H13, DNA
 DEFINITION

ACCESSION AQ027122
 VERSION AQ027122.1 GI:3267344
 KEYWORDS GSS.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 435)
 AUTHORS Adams.M.D., Rounsley.S.D., Zhao.S., Field.C.E., Bass.S., Linher.K.,
 Golden.K., Berry.K., Granger.D., Suh.E., Wible.C., Shizuya.H.,
 Simon.M. and Venter.J.C.
 TITLE Use of a random BAC End Sequence Database for Sequence-Ready Map
 Building (1998)

JOURNAL Unpublished (1998)
 COMMENT Contact: Mark Adams
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: mdadams@igrr.org

Clones are available from Research Genetics (info@resgen.com). BAC
 end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
 Seq primer: M13-21
 Class: BAC ends.

FEATURES

source
 1. .435
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="2323H13"
 /clone_lib="CIT-HSP"
 /sex="Male"

/cell_type="Sperm"
 /note="Vector: pBelobAC11; Site_1: HindIII; Site_2:
 HindIII"

BASE COUNT 140 a 97 c 77 g 121 t
 ORIGIN

Query Match 60.0%; Score 15; DB 223; Length 435;
 Best Local Similarity 100.0%; Pred. No. 42;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 catgccagcatttc 16
 |||||

Db 178 CATGCCAGCATTTC 192

RESULT 15
 H69200/c 437 bp mRNA EST 19-OCT-1995
 LOCUS YR95F10.r1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone
 DEFINITION IMAGE:213067 5', mRNA sequence.

ACCESSION H69200
 VERSION H69200.1 GI:1030526
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 437)
 AUTHORS Hillier.L., Lennon.G., Becker.M., Bonaldo.M.F., Chiapelli.B.,
 Chisoe.S., Dietrich.N., DuBuque.T., Favello.A., Gish.W., Hawkins
 M., Hultman.M., Kucaba.T., Lacy.M., Le.N., Mardis.E., Moore
 B., Morris.M., Parsons.J., Prange.C., Rifkin.L., Rohlfing.T.,
 Schellenberg.K., Soares.M.B., Tan.F., Thierry-Mieg.J., Trevaskis.E.,
 Underwood.K., Wohlmann.P., Waterston.R., Wilson.R. and Marra.M.
 TITLE Generation and analysis of 280,000 human expressed sequence tags
 Genome Res. 6 (9), 807-828 (1996)
 JOURNAL 97044478
 MEDLINE
 COMMENT

Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu

Insert Size: 1098
 High quality sequence stops: 300
 Source: IMAGE Consortium, LLNL
 This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert Length: 1098 Std Error: 0.00

Seq primer: M13RP1
 High quality sequence stop: 300.

FEATURES

source
 Location/Qualifiers
 1. .437
 /organism="Homo sapiens"
 /db_xref="GDB:3778371"
 /db_xref="taxon:9606"
 /clone="IMAGE:213067"
 /clone_lib="Soares fetal liver spleen INFLS"
 /sex="male"
 /dev_stage="20 week-post conception fetus"
 /lab_host="DH10B (ampicillin resistant)"

/note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
 with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
 1st strand cDNA was primed with a Pac I - oligo(dT) primer
 15' AACTGGAAGAAATTAATTAAGATCTTTTTTTTTTTT 3',
 double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Pac I and cloned into the Pac I
 and Eco RI sites of the modified pT7T3 vector. Library
 went through one round of normalization. Library
 constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 97 a 113 c 89 g 134 t
 ORIGIN

AV722315/c

LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AV722315 445 bp mRNA EST 16-OCT-2000
AV722315 HTB Homo sapiens cDNA clone HTBAMG01 5', mRNA sequence.
AV722315
AV722315.1 GI:10824678
EST.
human.
Homo sapiens

REFERENCE

AUTHORS
TITLE
JOURNAL
COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 445)
Gu, Y., Peng, Y., Song, H., Huang, Q., Yang, Y., Gao, G., Xiao, H., Xu, X.,
Li, N., Qian, B., Liu, F., Qu, J., Gao, X., Cheng, Z., Xu, Z., Zeng, L., Xu
S., Gu, W., Tu, Y., Jia, J., Fu, G., Ren, S., Zhong, M., Lu, G., Hu, R.,
Chen, J., Chen, Z., and Han, Z.
Homo sapiens cDNA HTB clones
Unpublished (2000)
Contact: Zequang Han
Chinese National Human Genome Center at Shanghai
351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
201203, P. R. China
Tel: 86-21-50801919(ex.45)
Fax: 86-21-50801922
Email: hanzg@chgc.sh.cn
This clone is available at CHGC in Shanghai.

FEATURES

source
Location/Qualifiers

1..445
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="HTBAMG01"
/clone_lib="HTB"
/tissue_type="Hypothalamus"
/dev_stage="Adult"
/lab_host="SOLR"
/note="Vector: pBluescript sk(-); Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT

ORIGIN

102 a 123 c 104 g 116 t
Query Match 60.0%; Score 15; DB 32; Length 445;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

Db

2 catcgccagcatttc 16
|||||
296 CATGCCGAGCATTC 282

RESULT 18

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

AA026650 466 bp mRNA EST 09-MAY-1997
zj99e08.r1 Soares_pregnant_uterus_NDHPU Homo sapiens cDNA clone
IMAGE:469190 5', mRNA sequence.
AA026650
AA026650.1 GI:1492715
EST.
human.
Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 466)
Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,
Chisoso, S., Dietrich, N., DuBuque, T., Favello, A., Gish, W., Hawkins
M., Mullman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Mardis, E., Moore
B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T.,
Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevaskis, E.,
Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Marra, M.
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
97044478
Contact: Wilson RK
Washington University School of Medicine

QY

Db

2 catcgccagcatttc 16
|||||

308 CATGCCGAGCATTC 294

RESULT 17

Query Match

Best Local Similarity

Matches

Conservative

Mismatches

Indels

Gaps

0; 0; 0; 0; 0; 0;

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu

This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1129 Std Error: 0.00
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 364.

FEATURES

Location/Qualifiers
1. 466
/organism="Homo sapiens"
/db_xref="GDB:375192"
/db_xref="taxon:9606"
/clone="IMAGE:469190"
/clone_lib="Soares_pregnant_uterus_NBHPU"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: uterus; Vector: pT7T3-Pac; Site_1: Not I;
Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer [5',
AACTGGAAGATTTCGCGCGCCCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization. Library
constructed by M. Fatima Bonaldo."

BASE COUNT 114 a 124 c 100 g 128 t
ORIGIN

Query Match 60.0%; Score 15; DB 1; Length 466;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 catgccagcatttc 16
|||||
Db 263 CATGCCAGCATTTC 249

RESULT 19

H60979/c 475 bp mRNA EST 06-OCT-1995
LOCUS
DEFINITION YR22f01.rl Soares fetal liver spleen INFLS Homo sapiens cDNA clone
IMAGE:206041 5', mRNA sequence.

ACCESSION H60979
VERSION H60979.1 GI:1013811
KEYWORDS EST.
SOURCE human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Hillier, L., Clark, N., Dubouche, T., Elliston, K., Hawkins, M., Holman,
M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J.,
Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston,
R., Williamson, A., Wohlmann, P., and Wilson, R.
The WashU-Merck EST Project
Unpublished (1995)
Contact: Willson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

TITLE

JOURNAL
COMMENT
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 1101
High quality sequence stops: 339
Source: IMAGE Consortium, LNL
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1101 Std Error: 0.00
Seq primer: M13RP1

High quality sequence stop: 339.
Location/Qualifiers
1. 475
/organism="Homo sapiens"
/db_xref="GDB:375172"
/db_xref="taxon:9606"
/clone="IMAGE:206041"
/clone_lib="Soares fetal liver spleen INFLS"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
1st strand cDNA was primed with a Pac I - oligo(dT) primer
[5' AACTGGAAGATTAAATGAATCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M. Fatima Bonaldo."

FEATURES

source

BASE COUNT 109 a 123 c 95 g 143 t 5 others
ORIGIN

Query Match 60.0%; Score 15; DB 158; Length 475;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 catgccagcatttc 16
|||||
Db 119 CATGCCAGCATTTC 105

RESULT 20

AI692369 491 bp mRNA EST 17-DEC-1999
LOCUS
DEFINITION wd63f02.x1 NCI_CGAP_Lu24 Homo sapiens cDNA clone IMAGE:2336283 3',
mRNA sequence.

ACCESSION AI692369
VERSION AI692369.1 GI:4969709
KEYWORDS EST.
SOURCE human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 491)

REFERENCE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

CONTACT: Robert Strausberg, Ph.D.

Email: cgapsb@mail.nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: M. Bento Soares, Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 589 Std Error: 0.00

Seq primer: -400P from Gibco

High quality sequence stop: 471.

Location/Qualifiers
1. 491

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:2336283"

/clone_lib="NCI_CGAP_Lu24"

/tissue_type="carcinoid"

/lab_host="DH10B"

/note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
modified polylinker; Plasmid DNA from the normalized

library NCI_CGAP_Lu5 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 1414920-1417991 and 1520904-1524339). Subtraction by Bento Soares and M. Fatima Bonaldo.

151 a 94 c 129 g 116 t 1 others

BASE COUNT
ORIGIN

Query Match 60.0%; Score 15; DB 23; Length 491;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 catcgccagcatttc 16
|||||
Db 344 CATGCCAGCATTC 358

RESULT 21

BF211097/c 491 bp mRNA EST 06-NOV-2000
LOCUS 601812706F1 NIH_MGC_54 Homo sapiens cDNA clone IMAGE:4047120 5',
DEFINITION mRNA sequence.

ACCESSION BF211097

VERSION BF211097.1 GI:11104683

KEYWORDS EST

SOURCE human.

ORGANISM

Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

NIH-MGC <http://mgc.nci.nih.gov/>.

1 (bases 1 to 491) National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapsb-r@mail.nih.gov

Tissue Procurement: ATCC

CNA Library Preparation: CLONETECH Laboratories, Inc.

CNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LNL at:

<http://image.lnl.gov>

Plate: LICM875 row: 1 column: 01

High quality sequence start: 56

High quality sequence stop: 491.

FEATURES

source

1..491
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4047120"
/clone_lib="NIH_MGC_54"
/tissue_type="from chronic myelogenous leukemia"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: bone marrow; Vector: pONR-LIB (Clontech);
Site_1: SfII (ggcgctcgcc); Site_2: SfII (ggccattatggcc
); Double-stranded cDNA was prepared from cell line RNA.
5' and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-CACGCCATTATGGC-3' and 3' adaptor
sequence: 5'-ATTCTAGCGCGGCGGCACATG-dh(30)BN-3',
(where B = A, C, or G and N = A, C, G, or T). Average
insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA)."

BASE COUNT 123 a 128 c 119 g 121 t

ORIGIN

Query Match 60.0%; Score 15; DB 146; Length 491;
Best Local Similarity 100.0%; Pred. No. 43;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 2 catcgccagcatttc 16
|||||
Db 305 CATGCCAGCATTC 291

RESULT 22

AZ924681/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

1..492

/organism="Saccharomyces mikatae"

/strain="IFO 1815"

/db_xref="taxon:114525"

/clone="4906.id55a24.sl"

/clone_lib="Saccharomyces mikatae IFO 1815"

/note="Random genomic sequence"

BASE COUNT 149 a 83 c 91 g 169 t

ORIGIN

Query Match 60.0%; Score 15; DB 251; Length 492;

Best Local Similarity 100.0%; Pred. No. 43;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 ccagcatttcgattcc 21

|||||

Db 475 CCAGCATTCGATCC 461

RESULT 23

T84279/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

Unpublished (1995)

Query Match 60.0%; Score 15; DB 146; Length 491;

Best Local Similarity 100.0%; Pred. No. 43;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 ccagcatttcgattcc 21

|||||

Db 475 CCAGCATTCGATCC 461

RESULT 23

T84279/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

Unpublished (1995)

AZ924681 492 bp DNA GSS 01-APR-2001
4906.id55a24.sl Saccharomyces mikatae IFO 1815 Saccharomyces
mikatae genomic clone 4906.id55a24.sl, DNA sequence.

AZ924681 492 bp DNA GSS 01-APR-2001
4906.id55a24.sl Saccharomyces mikatae IFO 1815 Saccharomyces
mikatae genomic clone 4906.id55a24.sl, DNA sequence.

AZ924681 492 bp DNA GSS 01-APR-2001
4906.id55a24.sl Saccharomyces mikatae IFO 1815 Saccharomyces
mikatae genomic clone 4906.id55a24.sl, DNA sequence.

AZ924681 492 bp DNA GSS 01-APR-2001
4906.id55a24.sl Saccharomyces mikatae IFO 1815 Saccharomyces
mikatae genomic clone 4906.id55a24.sl, DNA sequence.

AZ924681 492 bp DNA GSS 01-APR-2001
4906.id55a24.sl Saccharomyces mikatae IFO 1815 Saccharomyces
mikatae genomic clone 4906.id55a24.sl, DNA sequence.

AZ924681 492 bp DNA GSS 01-APR-2001
4906.id55a24.sl Saccharomyces mikatae IFO 1815 Saccharomyces
mikatae genomic clone 4906.id55a24.sl, DNA sequence.

AZ924681 492 bp DNA GSS 01-APR-2001
4906.id55a24.sl Saccharomyces mikatae IFO 1815 Saccharomyces
mikatae genomic clone 4906.id55a24.sl, DNA sequence.

AZ924681 492 bp DNA GSS 01-APR-2001
4906.id55a24.sl Saccharomyces mikatae IFO 1815 Saccharomyces
mikatae genomic clone 4906.id55a24.sl, DNA sequence.

AZ924681 492 bp DNA GSS 01-APR-2001
4906.id55a24.sl Saccharomyces mikatae IFO 1815 Saccharomyces
mikatae genomic clone 4906.id55a24.sl, DNA sequence.

AZ924681 492 bp DNA GSS 01-APR-2001
4906.id55a24.sl Saccharomyces mikatae IFO 1815 Saccharomyces
mikatae genomic clone 4906.id55a24.sl, DNA sequence.

AZ924681 492 bp DNA GSS 01-APR-2001
4906.id55a24.sl Saccharomyces mikatae IFO 1815 Saccharomyces
mikatae genomic clone 4906.id55a24.sl, DNA sequence.

AZ924681 492 bp DNA GSS 01-APR-2001
4906.id55a24.sl Saccharomyces mikatae IFO 1815 Saccharomyces
mikatae genomic clone 4906.id55a24.sl, DNA sequence.

AZ924681 492 bp DNA GSS 01-APR-2001
4906.id55a24.sl Saccharomyces mikatae IFO 1815 Saccharomyces
mikatae genomic clone 4906.id55a24.sl, DNA sequence.

AZ924681 492 bp DNA GSS 01-APR-2001
4906.id55a24.sl Saccharomyces mikatae IFO 1815 Saccharomyces
mikatae genomic clone 4906.id55a24.sl, DNA sequence.

AZ924681 492 bp DNA GSS 01-APR-2001
4906.id55a24.sl Saccharomyces mikatae IFO 1815 Saccharomyces
mikatae genomic clone 4906.id55a24.sl, DNA sequence.

AZ924681 492 bp DNA GSS 01-APR-2001
4906.id55a24.sl Saccharomyces mikatae IFO 1815 Saccharomyces
mikatae genomic clone 4906.id55a24.sl, DNA sequence.

AZ924681 492 bp DNA GSS 01-APR-2001
4906.id55a24.sl Saccharomyces mikatae IFO 1815 Saccharomyces
mikatae genomic clone 4906.id55a24.sl, DNA sequence.

AZ924681 492 bp DNA GSS 01-APR-2001
4906.id55a24.sl Saccharomyces mikatae IFO 1815 Saccharomyces
mikatae genomic clone 4906.id55a24.sl, DNA sequence.

AZ924681 492 bp DNA GSS 01-APR-2001
4906.id55a24.sl Saccharomyces mikatae IFO 1815 Saccharomyces
mikatae genomic clone 4906.id55a24.sl, DNA sequence.

AZ924681 492 bp DNA GSS 01-APR-2001
4906.id55a24.sl Saccharomyces mikatae IFO 1815 Saccharomyces
mikatae genomic clone 4906.id55a24.sl, DNA sequence.

AZ924681 492 bp DNA GSS 01-APR-2001
4906.id55a24.sl Saccharomyces mikatae IFO 1815 Saccharomyces
mikatae genomic clone 4906.id55a24.sl, DNA sequence.

AZ924681 492 bp DNA GSS 01-APR-2001
4906.id55a24.sl Saccharomyces mikatae IFO 1815 Saccharomyces
mikatae genomic clone 4906.id55a24.sl, DNA sequence.

AZ924681 492 bp DNA GSS 01-APR-2001
4906.id55a24.sl Saccharomyces mikatae IFO 1815 Saccharomyces
mikatae genomic clone 4906.id55a24.sl, DNA sequence.

Seq primer: puc 18 forward
 High quality sequence start: 8
 High quality sequence stop: 537.
 Location/Qualifiers
 1..539
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="HT0353"
 /dev_stage="Adult"
 /note="Organ: head_neck; Vector: puc18; Site_1: SmaI;
 Site_2: SmaI; A mini-library was made by cloning products
 derived from ORESSES PCR (U.S. Letters Patent application
 No. 196,716 - Ludwig Institute for Cancer Research)
 profiles into the pUC 18 vector. Reverse transcription of
 tissue mRNA and cDNA amplification were performed under
 low stringency conditions."
 BASE COUNT 124 a 143 c 128 g 143 t 1 others
 ORIGIN

Query Match 60.0%; Score 15; DB 170; Length 539;
 Best Local Similarity 100.0%; Pred. No. 43;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 catcgccagcatttc 16
 |||||||
 Db 400 CATGCCAGCATTC 386

RESULT 26
 AW588900/c
 LOCUS AW588900 547 bp mRNA EST 05-APR-2001
 DEFINITION ra06g05.y2 Bird-Rao Meloidogyne incognita J2 Meloidogyne incognita
 cDNA 5' similar to TR:055013 O55013 BET3. ;contains Alu repetitive
 element; , mRNA sequence.
 ACCESSION AW588900
 VERSION AW588900.1 GI:7275932
 KEYWORDS EST.
 SOURCE southern root-knot nematode.
 ORGANISM Meloidogyne incognita
 Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;
 Tylenchoidea; Heteroderidae; Meloidogyninae; Meloidogyne.
 1 (bases 1 to 547)
 McCarter,J., Clifton,S., Chiapelli,B., Pape,D., Martin,J., Wylie,T.,
 Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B., Bowers,Y.,
 Gibbons,M., Ritter,E., Bennett,J., Franklin,C., Tsagarishvili,R.,
 Ronko,I., Kennedy,S., Maguire,L., Beck,C., Underwood,K., Steptoe
 ,M., Allen,M., Person,B., Swaller,T., Harvey,N., Schurk,R., Kohn,S.,
 Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and
 Wilson,R.

TITLE The Washington Univ. Nematode EST Project, 1999
 JOURNAL Unpublished (1999)
 COMMENT Contact: McCarter JP
 The Washington Univ. Nematode EST Project, 1999
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 The library was constructed by Uma Rao and David Bird
 (david.bird@ucsu.edu) at North Carolina State University. DNA
 Sequencing by: Washington University Genome Sequencing Center St.
 Louis.
 Seq primer: T3 ET from Amersham
 High quality sequence stop: 519.
 Location/Qualifiers
 1..547
 /organism="Meloidogyne incognita"
 /db_xref="taxon:6306"
 /clone_lib="Bird-Rao Meloidogyne incognita J2"
 /dev_stage="enriched for 2nd stage juveniles"
 /lab_host="XLORL"
 /note="Vector: ZAP express - pBKCMV (Stratagene); Site_1:

FEATURES
 Source

EcoRI; Site_2: XhoI; Oligo (dT) primed library. cDNA was
 constructed and cloned unidirectionally into the vector
 within the 5' EcoRI and 3' XhoI sites. This library was
 constructed by Dr. Uma Rao and Dr. David Bird at North
 Carolina state University."
 BASE COUNT 179 a 76 c 106 g 185 t 1 others
 ORIGIN

Query Match 60.0%; Score 15; DB 118; Length 547;
 Best Local Similarity 100.0%; Pred. No. 43;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 catcgccagcatttc 16
 |||||||
 Db 296 CATGCCAGCATTC 282

RESULT 27
 AA405993/c
 LOCUS AA405993 581 bp mRNA EST 09-NOV-1997
 DEFINITION zu66f09.r1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:742985
 5' similar to contains MER29.b3 MER29 repetitive element ;, mRNA
 sequence.
 ACCESSION AA405993
 VERSION AA405993.1 GI:2063976
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 581)
 AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
 Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin
 ,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
 White,Y., Wylie,T., Waterston,R. and Wilson,R.
 WashU-NCI human EST Project
 Unpublished (1997)
 TITLE Contact: Wilson RK
 JOURNAL Washington University School of Medicine
 COMMENT 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert Length: 997 Std Error: 0.00
 Seq primer: -28ml3 rev2 ET from Amersham
 High quality sequence stop: 498.

FEATURES
 Source

1..581
 /organism="Homo sapiens"
 /db_xref="GDB:5930030"
 /db_xref="taxon:9606"
 /clone="IMAGE:742985"
 /clone_lib="Soares_testis_NHT"
 /sex="male"
 /lab_host="DH10B"
 /note="Vector: pf773D-pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
 was prepared from mRNA obtained from Clontech Laboratories
 , Inc., and primed with a Not I - oligo(dT) primer [5',
 TGTTACCAATCTGAAGTGGAGCGGCCCAATTTTTTTTTTTT 3'].
 Double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pf773 vector. Library
 went through one round of normalization to Cot5, and was
 constructed by Bento Soares and M. Fatima Bonaldo."
 BASE COUNT 153 a 171 c 131 g 126 t
 ORIGIN

Query Match

60.0%; Score 15; DB 7; Length 581;

```

Best Local Similarity 100.0%; Pred. No. 43;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 catcgccagcatttc 16
|||||
Db 81 CATCGCCAGCATTTC 67

RESULT 28
BE557935/c
LOCUS
DEFINITION
  fl15d06.y1 Zebrafish Research Genetics C32 fin Danio rerio cDNA 5'
  similar to TR:O18973 O18973 RAB5 GDP/GTP EXCHANGE FACTOR, RABEX5.
  ; mRNA sequence.
ACCESSION
  BE557935
VERSION
  BE557935.1 GI:9822425
KEYWORDS
  EST.
SOURCE
  zebrafish.
ORGANISM
  Danio rerio
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
  Cypriniformes; Cyprinidae; Rasbora; Danio.
REFERENCE
  1 (bases 1 to 622)
  Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
  ,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
  ,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
  Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
  Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
  and Wilson,R.
  WashU Zebrafish EST Project 1998
  Unpublished (1998)
  Contact: Stephen L. Johnson
  Washington University School of Medicine
  4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
  Tel: 314 286 1800
  Fax: 314 286 1810
  Email: zbrafish@watson.wustl.edu
  cDNA library preparation: Ning Wu. cDNA Library Arrayed by:
  Research Genetics. DNA Sequencing by: Washington University Genome
  Sequencing Center. Clone distribution: Research Genetics web
  address: http://www.researchgenetics.com/
  Seq primer: T3 ET from Amersham
  High quality sequence stop: 390.
FEATURES
  source
  1..622
  /db_xref="taxon:7955"
  /clone_lib="Zebrafish Research Genetics C32 fin"
  /tissue_type="Fin"
  /lab_host="GeneHogs (HS996, a phage-resistant isolate of
  DH10B)"
  /note="Vector: pT73D-Pac with a modified polylinker;
  Site_1: EcoRI; Site_2: NotI; 1st strand cDNA was prepared
  from zebrafish(C32) fin, and was then primed with a Not I
  - oligo(dT) primer. Double-stranded cDNA was ligated to
  Eco RI adaptors (Pharmacia), digested with Not I and
  cloned into the Not I and Eco RI sites of the modified
  pT73 vector. Library is non-normalized. Library was
  constructed by Ning Wu. NOTE:Clones from this library are
  only available thru Research Genetics (www.resgen.com)."
```

```

BASE COUNT 164 a 160 c 164 g 134 t
ORIGIN

Query Match 60.0%; Score 15; DB 137; Length 622;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 tcgcgcagcatttoga 18
|||||
Db 241 TCGCCAGCATTTCGA 227

RESULT 29
BE583417/c
LOCUS
DEFINITION
  11-6H-MY Psojaemy Phytophthora sojae cDNA, mRNA sequence.
ACCESSION
  BE583417
VERSION
  BE583417.1 GI:9834366
KEYWORDS
  EST.
SOURCE
  Phytophthora sojae.
ORGANISM
  Phytophthora sojae
  Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;
  Phytophthora.
  1 (bases 1 to 630)
  Qutob,D., Hraber,P.T., Sobral,B.W.S. and Gijzen,M.
  Comparative analysis of expressed sequences in Phytophthora sojae
  Plant Physiol. 123 (1), 243-254 (2000)
  20267956
  Contact: Gijzen M
  Agriculture and Agri-Food Canada
  1391 Sandford Street, London, Ontario, Canada N5V 4T3
  Tel: 519 457 1470
  Fax: 519 457 3997
  Email: gijzenm@em.agr.ca.
FEATURES
  source
  1..630
  Location/Qualifiers
  /organism="Phytophthora sojae"
  /strain="race 2, strain P6497"
  /db_xref="taxon:67593"
  /clone_lib="Psojaemy"
  /dev_stage="mycelium"
  /lab_host="E. coli strain XL0LR"
  /note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI; This
  cDNA library was constructed from polyA+ enriched mRNA
  from mycelium grown in liquid medium for 3 weeks at 25 C
  in the dark in a liquid synthetic medium of 2.4 g sucrose
  , 0.27 g asparagine, 10 mg cholesterol, 10 mg ascorbate,
  and 2 mg thiamine, per liter, plus salts and minerals.
  Complementary DNA was synthesized from mRNA using an
  XhoI-poly(dT) linker-primer. EcoRI adapters were ligated
  to the blunt-ended cDNA fragments and the products were
  digested with XhoI for directional cloning into lambda ZAP
  Express vector. This lambda library was amplified once
  using E. coli host strain XL1 Blue MRF'. Inserts were then
  subcloned by mass excision using ExAssist helper phage for
  conversion into phagenid vector pBK-CMV in E. coli host
  strain XL0LR. Sequenced using T3 primer: 5' ATT AAC CCT
  CAC TAA AGG GA 3'."
```

```

BASE COUNT 135 a 184 c 169 g 140 t
ORIGIN

Query Match 60.0%; Score 15; DB 137; Length 630;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 catcgccagcatttc 16
|||||
Db 16 CATCGCCAGCATTTC 2

RESULT 30
BE156891/c
LOCUS
DEFINITION
  fl158b12.y1 Sugano Kawakami zebrafish DRA Danio rerio cDNA clone
  3818062.5, similar to TR:Q9UJ41 Q9UJ41 RAB5 GDP/GTP EXCHANGE FACTOR
  HOMOLOGUE. ; mRNA sequence.
ACCESSION
  BE156891
VERSION
  BE156891.1 GI:11052079
KEYWORDS
  EST.
SOURCE
  zebrafish.
ORGANISM
  Danio rerio
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
  Cypriniformes; Cyprinidae; Rasbora; Danio.
```

Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabps@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM9278 row: o column: 21
High quality sequence stop: 630.
Location/Qualifiers
1. .1096
/organism="Mus musculus"
/strain="CZECH II"
/db_xref="taxon:10090"
/clone="IMAGE:4022252"
/clone_lib="NCI_CGAP_Lu30"
/tissue_type="tumor, metastatic to mammary"
/lab_host="DH10B"
/note="Organ: lung; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; transgenic model WNT-1, expression driven by MMTV-LTR enhancer; Cloned unidirectionally. Primer: Oligo dt. Library constructed by Life Technologies."
Investigator providing samples: Gilbert Smith, NIH"

336 a 271 c 271 g 218 t
h 60.0% Score 15; DB 145; Length 1096;
Similarity 100.0%; Pred No. 44;
15; Conservative 0; Mismatches 0; Indels 0; Gaps 0

attcgatctcgcg 24
|||||
ATTTCGATCTCG 1062

W06020 195 bp mRNA EST 22-MAY-2000
TGESTy87f10.r1 TqR tachyzoite CDNA Toxoplasma gondii cDNA clone
tazy87f10.r1 5', mRNA sequence.

W06020
W06020.1 GI:1278733
EST.
Toxoplasma gondii.
Toxoplasma gondii
Eukaryota; Alveolata; Apicomplexa; Coccidia; Elmeriida;
Sarcocystidae; Toxoplasma.
1 (bases 1 to 195)
Hehl,A., Manger,I., Marra,M., Sibley,L.D., Ajioke,J.A., Aslett,M.A.,
Dietrich,N., Dubouque,T., Hillier,D., Kucaba,T., Wan,K.L.,
Waterston,R.H. and Boothroyd,J.
Washington, R.H. and Boothroyd, J.
WashU-Merck-Stanford-NIH Toxoplasma EST project
Unpublished (1996)
Contact: Marra M
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: toxo@watson.wustl.edu
David Sibley at toxest@borcim.wustl.edu for further information
relating to organism, clone or library availability.
Seq primer: T3
High quality sequence stop: 148.
Location/Qualifiers
1. .195
/organism="Toxoplasma gondii"
/strain="RH"
/db_xref="taxon:5811"

1 (bases 1 to 684)
 Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
 ,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
 ,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
 Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
 Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
 and Wilson,R.
 WashU Zebrafish EST Project 1998
 Unpublished (1998)
 Contact: Stephen L. Johnson
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: zbrafish@watson.wustl.edu
 Library constructed by Dr. Sumio Sugano and Dr. Koichi Kawakami DNA
 Sequencing by: Washington University Genome Sequencing Center Clome
 distribution information can be found through the I.M.A.G.E.
 Consortium/LLNL, send email to: info@image.llnl.gov
 Seq primer: T3 ET from Amersham
 High quality sequence stop: 501.

Location/Qualifiers
 1..684
 /organism="Danio rerio"
 /strain="AB"
 /db_xref="taxon:7955"
 /clone="3818062"
 /clone_lib="Sugano Kawakami zebrafish DRA"
 /sex="mixed (one male and one female, including
 unfertilized eggs)"
 /dev_stage="adult"
 /lab_host="DH10B (phage resistant)"
 /note="Vector: pME18S-FL3; Site.1: DraIII (CACTGTGTG);
 Site.2: DraIII (CACCATGTG); 1st strand cDNA was primed
 with an oligo(dT) primer [ATGTGCGCTTTTTTTTTTTTTT];
 double-stranded cDNA was ligated to a DraIII adaptor
 [TGTGTCCTACTGTG], digested and cloned into distinct DraIII
 sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site
 CACCATGTG). XhoI should be used to isolate the cDNA
 insert. Size selection was performed to exclude fragments
 <1.5kb. Library constructed by Dr. Sumio Sugano
 (University of Tokyo Institute of Medical Science) and
 kindly donated by Dr. Koichi Kawakami. Custom primers for
 sequencing: 5' end primer CTTCCTGCTCTAAAGCTGCG and 3' end
 primer CGACCTCGACCTCGACACA."

BASE COUNT 177 a 177 c 193 g 137 t
 ORIGIN

Query Match 60.0%; Score 15; DB 145; Length 684;
 Best Local Similarity 100.0%; Pred. No. 43;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 tcgccagcatttcga 18
 |||||
 Db 237 TCCCAGCATTTCGA 223

RESULT 31
 BF144032/c
 LOCUS
 DEFINITION
 mRNA sequence.
 601791288F1 NCI_CGAP_Lu30 Mus musculus cDNA clone IMAGE:402252 5',
 BF144032
 BF144032.1 GI:10983072
 EST.
 EST.
 house mouse.
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 1096)
 NIH-MGC http://mgc.nci.nih.gov/.
 National Institutes of Health, Mammalian Gene Collection (MGC)

REFERENCE
 AUTHORS
 TITLE

FEATURES
 SOURCE

BASE COUNT
 ORIGIN

Query Match
 Best Local
 Matches

QY 10
 Db 1076

RESULT 32
 W06020/c
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

REFERENCE
 AUTHORS

TITLE
 JOURNAL
 COMMENT

FEATURES
 SOURCE

/clone="tgzy87f10.x1"
 /clone_lib="TGRH Tachyzoite cDNA"
 /lab_host="XLI-Blue MRP"
 /note="Vector: Lambda ZAP; Site_1: EcoRI; Site_2: XhoI;
 Toxoplasma RH strain tachyzoites were grown in human
 foreskin fibroblast cultures in vitro. The library was
 constructed by K.L. Wan, Cambridge University. cDNAs were
 synthesized from polyA RNAs by oligo d(T) priming and
 directionally cloned into the EcoRI to XhoI sites of the
 Lambda ZapII vector using the ZAP-cDNA synthesis kit
 (Stratagene). WARNING: the library contains a small
 percentage of cDNAs derived from the human host cells."
 59 a 45 c 62 g 27 t 2 others

BASE COUNT 59 a 45 c 62 g 27 t 2 others
 ORIGIN

Query Match 56.0%; Score 14; DB 189; Length 195;
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 gcatttcgacctc 23
 |||||

Db 181 GCATTTCGATCCTC 168

RESULT 33

AZ922752 200 bp DNA GSS 22-MAR-2001
 LOCUS SLCot4A05 Sorghum bicolor SLCot Sorghum bicolor genomic, DNA
 DEFINITION

ACCESSION AZ922752
 VERSION AZ922752.1 GI:13432973
 KEYWORDS GSS.
 SOURCE sorghum.

ORGANISM

Sorghum bicolor
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade; Panicoideae; Andropogoneae; Sorghum.

1 (bases 1 to 200)
 Peterson,D.G., Schulze,S.R., Lee,S.A., Sciarra,E.B., Nagel,A.,
 Tibbitts,D.C., Wessler,S.R. and Paterson,A.H.

Tibbitts,D.C., Wessler,S.R. and Paterson,A.H.
 Characterization of the Sorghum bicolor genome using DNA

renaturation kinetics (Cot analysis) and repetition-based cloning

Unpublished (2001)

Contact: Peterson DG

Plant Genome Mapping Laboratory

University of Georgia

Room 162, Riverbend Research Bldg., 110 Riverbend Rd., Athens, GA

30602 USA

Tel.: 706-583-0167

Fax: 706-583-0160

Email: dgp@arches.uga.edu

Class: Hydroxyapatite-fractionated DNA.

Location/Qualifiers

1..200

/organism="Sorghum bicolor"

/cultivar="Btx623"

/db_xref="taxon:4558"

/clone_lib="Sorghum bicolor SLCot"

/tissue_type="leaves"

/dev_stage="seedling"

/note="Vector: pGEM-TA-Easy; A Cot analysis was performed

for the sorghum genome. Based on the resulting Cot curve,

hydroxyapatite chromatography was used to isolate

'highly-repetitive' (HR), 'moderately-repetitive' (MR),

and 'single/low-copy' (SL) sequence components from

sheared genomic DNA. The three repetition-based DNA

components were cloned into E. coli to produce HRCot,

MRCot, and SLCot genomic libraries. Blotting and

sequencing data indicates that each library is

representative of the component from which it was derived.

Putative ID listings given for sequences are based on

, comparison (blastn) with sequences in the NCBI Nr

Database. Only the primary match is given (all primary E
 values are < or =3D 1.00E-5). In no instance does a 'Cot
 clone' contain the complete sequence of its putative Nr
 match."

BASE COUNT 48 a 49 c 57 g 42 t 4 others
 ORIGIN

Query Match 56.0%; Score 14; DB 251; Length 200;
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 catgccagcattt 15
 |||||

Db 91 CATGCCAGCATT 104

RESULT 34

AW065325/c

LOCUS

DEFINITION

614045H11.y1 614 - root cDNA library from Walbot Lab Zea mays cDNA,

mRNA sequence.

ACCESSION AW065325

VERSION AW065325.1 GI:6020397

KEYWORDS EST.

SOURCE Zea mays.

ORGANISM Zea mays.

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC

clade; Panicoideae; Andropogoneae; Zea.

1 (bases 1 to 203)

Walbot,V.

Maize ESTs from various cDNA libraries sequenced at Stanford

University

Unpublished (1999)

Contact: Walbot V

Department of Biological Sciences

Stanford University

855 California Ave, Palo Alto, CA 94304, USA

Tel: 650 723 2227

Fax: 650 725 8221

Email: walbot@stanford.edu

Plate: 614045 row: H column: 11.

Location/Qualifiers

1..203

/organism="Zea mays"

/cultivar="W23"

/db_xref="taxon:4577"

/clone_lib="614 - root cDNA library from Walbot Lab"

/tissue_type="root"

/dev_stage="3-4 days old"

/lab_host="XLOLR"

/note="Organ: root; Vector: pBlueScriptII SK+; Site_1:

EcoRI; Site_2: XhoI; 3-4 days old root tissue from Walbot

Lab (LM)"

BASE COUNT 46 a 52 c 51 g 54 t

ORIGIN

Query Match 56.0%; Score 14; DB 111; Length 203;

Best Local Similarity 100.0%; Pred. No. 1.7e+02;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 atgcgcagcatttc 16
 |||||

Db 63 ATGCAGCATTTC 50

RESULT 35

BG589383/c

LOCUS

DEFINITION

ES497225 P. infestans-challenged leaf Solanum tuberosum cDNA clone

BPL11H7 5' sequence, mRNA sequence.

```

ACCESSION      BG589383
VERSION        BG589383.1  GI:13607523
KEYWORDS       EST.
SOURCE         SOLANUM
ORGANISM       Solanum tuberosum
               Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
               Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
               Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
REFERENCE      1 (bases 1 to 203)
AUTHORS        Zhang,P., Hernandez,M., Torndqvist,C.-E., Wirtz,U., Loukolanov,A.,
               Rangell,P., Haberlandt,G.T., Cho,J., Chiemingo,A., Bougri,O., Buell
               ,C.R., Ronning,C.M., Helgeson,J. and Baker,B.
               Generation of ESTs from Potato Leaves Challenged with Phytophthora
               infestans, Incompatible Reaction
               Unpublished (2000)
JOURNAL         The Institute for Genomic Research
COMMENT         For clone info: please contact Research Genetics, Libraries
               Division tel 1-800-711-6195, email cdna@resgen.com
               Seq primer: M3F-R.
FEATURES       Location/Qualifiers
               1..203
               /organism="Solanum tuberosum"
               /cultivar="Kennebec"
               /db_xref="taxon:4113"
               /clone="BP111H7"
               /clone_lib="P. infestans-challenged leaf"
               /tissue_type="leaf"
               /dev_stage="6 week old"
               /lab_host="SOLR"
               /note="vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
               XhoI; Whole plants were challenged with 450,000
               sporangia/ml P. infestans US-1(US 940501) in Biotron
               (Madison, Wisconsin). Leaf tissue was collected at 1, 2,
               5, 12, and 24 hours post-challenge and frozen in liquid
               nitrogen immediately upon removal. Kennebec plants showed
               no signs of HR. Katahdin plants (susceptible to P.
               infestans US-1) were used as controls and showed
               infection. NOTE: We cannot exclude the possibility that
               this sequence is actually derived from Phytophthora rather
               than potato."
BASE COUNT     67 a      68 g      30 t
ORIGIN
Query Match    56.0%; Score 14; DB 155; Length 203;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 tcgcacgacatttcg 17
Db 93 TCGCCAGCATTTCCG 80

RESULT 36
BG558891
LOCUS          RH122_57_H09.bl_A003 Rhizome2 (RH122) Sorghum propinquum cDNA, mRNA
DEFINITION     sequence.
ACCESSION      BG558891
VERSION        BG558891.1  GI:13587889
KEYWORDS       EST.
SOURCE         Sorghum propinquum.
ORGANISM       Sorghum propinquum
               Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
               Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
               clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE      1 (bases 1 to 205)
AUTHORS        Cordonnier-Pratt,M.-M., Gingle,A., Paterson,A., Sudman,M. and Pratt
               ,L.H.
               An EST database from Sorghum: Sorghum propinquum rhizomes
               Unpublished (2000)
JOURNAL         Contact: Cordonnier-Pratt MM

ACCESSION      BG589383
VERSION        BG589383.1  GI:13607523
KEYWORDS       EST.
SOURCE         SOLANUM
ORGANISM       Solanum tuberosum
               Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
               Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
               Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
REFERENCE      1 (bases 1 to 203)
AUTHORS        Zhang,P., Hernandez,M., Torndqvist,C.-E., Wirtz,U., Loukolanov,A.,
               Rangell,P., Haberlandt,G.T., Cho,J., Chiemingo,A., Bougri,O., Buell
               ,C.R., Ronning,C.M., Helgeson,J. and Baker,B.
               Generation of ESTs from Potato Leaves Challenged with Phytophthora
               infestans, Incompatible Reaction
               Unpublished (2000)
JOURNAL         The Institute for Genomic Research
COMMENT         For clone info: please contact Research Genetics, Libraries
               Division tel 1-800-711-6195, email cdna@resgen.com
               Seq primer: M3F-R.
FEATURES       Location/Qualifiers
               1..203
               /organism="Solanum tuberosum"
               /cultivar="Kennebec"
               /db_xref="taxon:4113"
               /clone="BP111H7"
               /clone_lib="P. infestans-challenged leaf"
               /tissue_type="leaf"
               /dev_stage="6 week old"
               /lab_host="SOLR"
               /note="vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
               XhoI; Whole plants were challenged with 450,000
               sporangia/ml P. infestans US-1(US 940501) in Biotron
               (Madison, Wisconsin). Leaf tissue was collected at 1, 2,
               5, 12, and 24 hours post-challenge and frozen in liquid
               nitrogen immediately upon removal. Kennebec plants showed
               no signs of HR. Katahdin plants (susceptible to P.
               infestans US-1) were used as controls and showed
               infection. NOTE: We cannot exclude the possibility that
               this sequence is actually derived from Phytophthora rather
               than potato."
BASE COUNT     67 a      68 g      30 t
ORIGIN
Query Match    56.0%; Score 14; DB 155; Length 205;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 atccacgacatttc 16
Db 141 ATGCCAGCATTTTC 154

RESULT 37
AQ931496
LOCUS          A0931496.1  GI:6620510
DEFINITION     DNA sequence.
ACCESSION      AQ931496
VERSION        AQ931496.1  GI:6620510
KEYWORDS       GSS.
SOURCE         house musculus.
ORGANISM       Mus musculus
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE      1 (bases 1 to 211)
AUTHORS        Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
               ,B., Levins,M., Mcgann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
               and Fraser,C.M.
               Mouse BAC End Sequences from Library RPCI-23
               Unpublished (1999)
TITLE          Contact: Shaying Zhao
JOURNAL        Department of Eukaryotic Genomics
COMMENT        The Institute for Genomic Research
               9712 Medical Center Dr., Rockville, MD 20850, USA
               Tel: 301 838 0200
               Fax: 301 838 0208
               Email: szhao@tigr.org
               Clones are derived from the mouse BAC library RPCI-23. For BAC
               library availability, please contact Pieter de Jong
               (pieter@dejong.med.buffalo.edu). Clones may be purchased from
               BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
               or from Resea ch Genetics (info@resgen.com). BAC end page:
               http://www.tigr.org/tdb/bac-ends/mouse/bac_end_intro.html
               Plate: 266 row: F column: 9
               Seq primer: SP6
               Class: BAC ends.
FEATURES       Location/Qualifiers
               1..211
               /organism="Mus musculus"
               /strain="C57BL/6J"
               /db_xref="taxon:10090"
               /clone="RPCI-23-266f9"
               /clone_lib="RPCI-23"

```

Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmpratt@uga.edu
Seq primer: JEN REV
High quality sequence start: 4
High quality sequence stop: 167
POLYA=No.

FEATURES

source

Location/Qualifiers

1..205

/organism="Sorghum propinquum"

/db_xref="taxon:132711"

/clone_lib="Rhizome2 (RH122)"

/note="Organ: Rhizomes; Vector: pBluescript II from Lambda

Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made

from poly-A RNA in the cloning vector lambda ZAP II.

Clones to be sequenced were prepared by mass excision."

BASE COUNT 60 a 45 c 52 g 48 t

ORIGIN

Query Match

56.0%; Score 14; DB 155; Length 205;

Best Local Similarity 100.0%; Pred. No. 1.7e+02;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 atccacgacatttc 16

Db 141 ATGCCAGCATTTTC 154

RESULT 37

AQ931496

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

AUTHORS

1 (bases 1 to 211)

Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret

,B., Levins,M., Mcgann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.

and Fraser,C.M.

Mouse BAC End Sequences from Library RPCI-23

Unpublished (1999)

TITLE

JOURNAL

COMMENT

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC

library availability, please contact Pieter de Jong

(pieter@dejong.med.buffalo.edu). Clones may be purchased from

BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)

or from Resea ch Genetics (info@resgen.com). BAC end page:

http://www.tigr.org/tdb/bac-ends/mouse/bac_end_intro.html

Plate: 266 row: F column: 9

Seq primer: SP6

Class: BAC ends.

FEATURES

source

Location/Qualifiers

1..211

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="RPCI-23-266f9"

/clone_lib="RPCI-23"

/sex="Female"
/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI Methylase. Size
selected DNA was cloned into the pBACe3.6 vector at the
EcoRI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT 82 a 38 c 28 g 62 t 1 others
ORIGIN

Query Match 56.0%; Score 14; DB 235; Length 211;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 6 gccagcattcgat 19
|||||
Db 159 GCCAGCATTTCGAT 172

RESULT 38
AI329089/c
LOCUS 231 bp mRNA EST 28-DEC-1998
DEFINITION a9l2ne.r1 Neurospora crassa evening cDNA library Neurospora crassa
CDNA clone a9l2ne 3', mRNA sequence.
ACCESSION AI329089
VERSION AI329089.1 GI:4065648
KEYWORDS EST.
SOURCE Neurospora crassa.
ORGANISM Neurospora crassa

REFERENCE
AUTHORS Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Sordariales; Sordariaceae; Neurospora.
TITLE 1 (bases 1 to 231)
JOURNAL Zhu, H., Lai, H., Kupfer, D., Dunlap, J.C. and Roe, B.A.
COMMENT Two Neurospora crassa EST Databases
Unpublished (1998)
Other ESTs: a9l2ne.fl

Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu
Department of Chemistry and Biochemistry
Advanced Center for Genome Technology, University of Oklahoma
620 Parrington Oval, Norman, OK 73019, USA
Tel: 405 325 4912
Fax: 405 325 7762
Email: broe@ou.edu
We anticipate the future release of the cDNA clones to the Fungal
Genetics Stock Center
Seq primer: Universal Reverse Primer
High quality sequence stop: 179.

FEATURES
source
1..231
/organism="Neurospora crassa"
/strain="Strain 30-7 (bd; A)"
/db_xref="taxon:5141"
/clone="a9l2ne"
/clone_lib="Neurospora crassa evening cDNA library"
/tissue_type="tissue harvested following 22hr growth in
dark"
/note="Vector: pBluescript SK-; Site_1: XbaI; Site_2:
EcoRI; See: Bell-Perdersen, D., et al. PNAS 93:13096, 1996.
5' end of cDNA cloned into XbaI site of pBluescript; 3'
end of cDNA cloned into EcoRI site of pBluescript."
BASE COUNT 50 a 71 c 39 g 71 t
ORIGIN

Query Match 56.0%; Score 14; DB 18; Length 231;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 gccatgcagcatt 14
|||||

Db 57 GCATCGCCAGCATT 44
RESULT 39
AI902416/c
LOCUS 242 bp mRNA EST 30-MAR-2000
DEFINITION CM-BT006-050299-125 BT006 Homo sapiens cDNA, mRNA sequence.
ACCESSION AI902416
VERSION AI902416.1 GI:6492803
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 242)
AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,
Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
M.J., Soares, F., Brentani, R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/seq/gethtml.pl?tl=CM&t2=CM-BT006-125.html
&t3=050299&t4=1)
Seq primer: puc 18 forward.

FEATURES
source
1..242
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="BT006"
/sex="female"
/dev_stage="Adult"
/note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
7,716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."
BASE COUNT 61 a 59 c 71 g 35 t 16 others
ORIGIN

Query Match 56.0%; Score 14; DB 103; Length 242;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 6 gccacattcgat 19
|||||

Db 219 GCCAGCATTTCGAT 206

RESULT 40
BB090429
LOCUS 251 bp mRNA EST 23-JUN-2000
DEFINITION BB090429 RIKEN full-length enriched, 12 days embryo, embryonic body
between diaphragm region and neck Mus musculus cDNA clone
9430026F22 3', similar to AL080159 Homo sapiens mRNA; cDNA
DKFZp434M154 (from clone DKFZp434M154);, mRNA sequence.
ACCESSION BB090429

BASE COUNT

ORIGIN

Query Match 56.0%; Score 14; DB 115; Length 271;
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 ccagcattcgatc 20
 |||||

Db 264 CCAGCATTTCGATC 251
 |||||

RESULT 42

AA165335
 LOCUS 286 bp mRNA EST 18-DEC-1996
 DEFINITION zq49c06.r1 Stratagene hNT neuron (#937233) Homo sapiens cDNA clone
 IMAGE:633034 5', mRNA sequence.

ACCESSION AA165335
 VERSION AA165335.1 GI:1741484
 KEYWORDS EST.
 SOURCE human.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 286)
 Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,
 Chisoe, S., Dietrich, N., DuBuque, T., Favell, A., Gish, W., Hawkins,
 M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Mardis, E., Moore,
 B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T.,
 Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevaskis, E.,
 Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Marra, M.
 Generation and analysis of 280,000 human expressed sequence tags
 Genome Res. 6 (9), 807-828 (1996)
 9704478

TITLE

CONTACT: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: -28M13 rev2 from Amersham

High quality sequence stop: 178.

FEATURES

source

Location/Qualifiers
 1..286
 /organism="Homo sapiens"
 /db_xref="GDB:5183118"
 /db_xref="taxon:9606"
 /clone="IWAGE:633034"
 /clone_lib="Stratagene hNT neuron (#937233)"
 /dev_stage="hNT neurons"
 /lab_host="SOLR (kanamycin resistant)"
 /note="vector: pBluescript SK-; Site_1: EcoRI; Site_2:
 XhoI; Cloned unidirectionally. Primer: Oligo dt.
 Differentiated, post mitotic hNT neurons. Average insert
 size: 1.5 kb; Uni-ZAP XR Vector: -5' adaptor sequence: 5'
 GAATTCGGCAGAG 3' -3' adaptor sequence: 5'
 CTCGAGTTTTTTTTTTTTTTTTTTT 3'

BASE COUNT

55 a 120 c 66 g 38 t 7 others

ORIGIN

Query Match 56.0%; Score 14; DB 3; Length 286;
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 gcacgcagcatt 14
 |||||

Db 241 GCATCGCAGCATT 254
 |||||

RESULT 43

BB216989

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

BB216989

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

BB216989

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

BB216989

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

BB216989

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

BB216989

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

BB216989

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

